

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 6, 2004, 17:52:00 ; Search time 43.5 Seconds
(without alignments)
7050.200 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876

Sequence: 1 gcacgaggtatnaaaat.....aaagggaantccccatggg 486

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2.1/USPTO.spool.p/US10057510/runat_06042004_141608_14722/app_query.fasta_1.647
-DB=SPTREMBL_25 -QWMT=Eastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10057510@cgn.1.1.86@runat_06042004_141608_14722 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plan.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	100.5	12.0	132	11 Q8BRC3	Q8brc3 mus musculus

C	2	98	11.7	258	12	Q917X0
C 3	94.5	11.2	633	10	Q7X188	
C 4	92	11.0	191	4	Q99491	
C 5	92	11.0	258	12	Q917W7	
C 6	91	10.8	495	5	Q8MSC6	
C 7	91	10.8	1355	5	Q9VKH2	
C 8	91	10.8	3469	5	Q9U412	
C 9	90.5	10.8	209	12	Q69115	
C 10	90.5	10.8	209	12	Q69123	
C 11	89.5	10.7	647	16	Q8XZN8	
C 12	89.5	10.7	2414	4	Q9HCL7	
C 13	88.5	10.5	1093	5	Q9V6T2	
C 14	88.5	10.5	1150	3	Q872W5	
C 15	86	10.2	378	16	Q8ZL04	
C 16	86	9.8	805	13	Q9PTY3	
C 17	85.5	10.2	1325	5	Q9BKV7	
C 18	85.5	10.2	1405	2	Q8XMK1	
C 19	84.5	10.1	439	10	Q7XEA2	
C 20	84.5	10.1	1173	11	Q63624	
C 21	84.5	10.1	1427	2	Q8GIZ2	
C 22	84.5	10.1	2957	5	Q61845	
C 23	84	9.6	336	4	Q7Z3T3	
C 24	84	10.0	381	11	Q8C096	
C 25	84	10.0	429	13	Q7ZWV6	
C 26	84	9.6	696	16	Q7VY84	
C 27	83.5	9.9	274	10	Q9M4M5	
C 28	83.5	9.5	336	4	Q81WP5	
C 29	83.5	9.9	383	5	Q9Y075	
C 30	83.5	9.9	817	3	Q07229	
C 31	83.5	9.9	3604	5	Q9VYK0	
C 32	83.5	9.9	230	4	Q96D64	
C 33	83	9.9	470	4	Q81YD6	
C 34	83	9.9	1706	16	Q9RZJ2	
C 35	83	9.9	1791	5	Q9U6D4	
C 36	82.5	9.8	325	10	Q9ZV88	
C 37	82.5	9.8	545	3	Q9HEK1	
C 38	82.5	9.8	749	10	C04387	
C 39	82.5	9.8	873	5	Q9Y076	
C 40	82.5	9.4	283	2	O31397	
C 41	82	9.4	479	16	Q9RKZ0	
C 42	82	9.8	811	16	Q7ZWS6	
C 43	81.5	9.7	414	10	Q94AS1	
C 44	81.5	9.3	478	12	Q91TL3	
C 45						

ALIGNMENTS

RESULT 1
Q8BRC3
ID Q8BRC3 PRELIMINARY; PRT; 132 AA.
AC Q8BRC3; 01-MAR-2003 (T-REMBLrel. 23, Created)
DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-REMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12456851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK045144; BAC32238.1; -.
KW Hypothetical protein.
SQ SEQUENCE 132 AA; 13680 MW; FD20A00CA4B554C CRC64;

Alignment Scores:		Length:	132
Pred. No.:	0.0548	Matches:	43
Score:	100.50	Conservative:	6
Percent Similarity:	33.33%	Mismatches:	56
Best Local Similarity:	29.25%	Indels:	42
Query Match:	11.96%	Gaps:	6
DB:	11		
US-10-057-510-1 (1-486) x Q8BRC3 (1-132)			
QY	473	TCCTTTCTTCAGGCTTGCNAAAACCGGGGNCNTANGNCCAGTTTTCAGAACAGA	414
DB	10	SerLeuArgSerGlyValGlyLeuSerProSerHisSerArgProArgValSerProArg	29
QY	413	GGCCAGGACACACCA-----GGGAGGGGCTTCTTATGCAAGTTCCCTTCAT	363
DB	30	GlyProGlyProAlaProProLysSerGlyArg-----	40
QY	362	GCCATATGAGCCAAAGNCTNGTGGCTCGAACCCCTTGACTGGGCAATGTCTCGAGACTGTG	303
DB	41	-----ProGlyAlaLeuProAlaSerLeu	48
QY	302	GGGACNATCCAGGGGTGCATCTTCATCTCAGCCCTGGCTA-----	261
DB	49	AlaValProPro-----ProProSerCysLeuTrpProGluArgGlnGln	63
QY	260	GCCTTCTGCCCTCTCCAGGGCTGTCATCTTCAGCTCCAGGCGCTCTCTCCACAGC	201
DB	64	ArgLeuAlaProArgProSerThrArgThrPheLeuAsnValProSerArgAla---Ala	82
QY	200	TCCATGCTCCACCTTGCACATGGANACAGCGGTGCCCCACACAGCAGGGAACAA	141
DB	83	SerLeuProAspAspLeuCysSerGlyThrSerSerAlaProLeuSer-ArgGlyLysAl	102
QY	140	CGAAGCT-----GCCAAGAAAGAACAGCAGTGTAGAGCAGGGGCAACAATG	96
DB	102	aArgAlaAspGlyArgGlySerGlyProProGluProGlnMetValSerMetTrpGlnTr	122
QY	95	GAGAAAGAGTATCTATTA	77
DB	122	pArgLysGlnAlaAlaLeu	128
RESULT 2			
ID	Q917X0	PRELIMINARY;	PRT; 258 AA.
AC	Q917X0;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Minor structural protein.		
GN	ORF3.		
OS	Norwalk-like virus NLV/New Orleans/306/1994/US.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;		
OC	Norovirus.		
OX	NCBI_TaxID=171854;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Hu/NLV/New Orleans/306/1994/US;		
RX	MEDLINE=97193806; PubMed=9041391;		
RA	Ando T., Monroe S.S., Noel J.S., Glass R.I.;		
RT	"A one-tube method of reverse transcription-PCR to efficiently amplify		
RT	a 3-kilobase region from the RNA polymerase gene to the poly(A) tail		
RT	of small round-structured viruses (Norwalk-like viruses).";		
RL	J. Clin. Microbiol. 35:570-577(1997).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Hu/NLV/New Orleans/306/1994/US;		
RX	MEDLINE=20266071; PubMed=10804147;		
RA	Ando T., Noel J.S., Fankhauser R.I.;		
RT	"Genetic classification of 'Norwalk-like viruses.'";		
RL	J. Infect. Dis. 181:S336-S348(2000).		
RP	[3]		
RP	SEQUENCE FROM N.A.		
STRAIN=Hu/NLV/New Orleans/306/1994/US;			
RA	Ando T., Seto Y., Noel J.S., Monroe S.S., Glass R.I., Fankhauser R.I.;		
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF414422; AAL13020.1; -		
DR	InterPro; IPR004278; RNA_capsid.		
DR	Pfam; PF03035; RNA_capsid; 1.		
SQ	SEQUENCE 258 AA; 27412 MW; 272E2A8FA8EA2A4 CRC64;		
Alignment Scores:			
Pred. No.:	0.112	Length:	258
Score:	98.00	Matches:	41
Percent Similarity:	40.49%	Conservative:	25
Best Local Similarity:	25.15%	Mismatches:	64
Query Match:	11.67%	Indels:	33
DB:	12	Gaps:	4
US-10-057-510-1 (1-486) x Q917X0 (1-258)			
QY	447	CGAGGGGNCNTANGNCCAGTTTTCAGAACAGAGGGCCAGCACACACCGAGGGG	388
DB	108	ProGlySerMet-----LysThrThrSerTySerGly	118
QY	387	GCTTCTTATGCAAGTTCCTTCATGTCATATGAGCCAAAGNCCTNGTGGCTCGAACCCCT	328
DB	119	LysPheValSerMetAsnPro-----ValArgGlnVal	129
QY	327	GACTGGGCAATGCTGAGAGCTGGGGACNATCCAGGGGTGCATCTTCATCTCCAGCC	268
DB	130	GluPheProGlnProLysLysSerAlaAlaIleProSerSerAlaSerValSerSer	149
QY	267	CTGCTAGCTTCTGCCCTCTCCAGGGCTCGTCCATTCCTTCAGCTCCAGCCCTCTCT	208
DB	150	GlyArgThrAsnLeuThrAsnSerThrGlnSerThrValSerThrSerSerAlaPro	169
QY	207	CCACAGC-----TCCATGCTTCCACCTTGCACATGGANACAGCCCTGCCCCAC	157
DB	170	Pro-SerArgGlySerSerAlaProSerThrLeuSerArgAlaThrThrArgThrSerAs	189
QY	156	RAGCCAGAGGGAACAAAGCTGCGAAGAAAGA-----	123
DB	189	nTrpValGluGlnAsnArgAsnLeuGluProTyMetArgGlyAlaLeuGlnThrTh	209
QY	122	-----ACAGCAGTTAGAGCAGGGGCAACATGGAGAAAGAGTATCTATATGAGCTT	70
DB	209	rPheValThrProProSerSerArgAlaSerSerAsnGlyThrValSerThrValProLy	229
QY	69	GGCATTTTACACGACTTCCAGCCCTTTCATAAATGGAACATTTTATATA	10
DB	229	sGlyValLeuAspSerTrpThrProAlaPheAsnThrArgArgGlnProLeuPheAlaTy	249
QY	9	CGCTCGT	3
DB	249	rLeuArg	251
RESULT 3			
ID	Q7X188	PRELIMINARY;	PRT; 633 AA.
AC	Q7X188;		
DT	01-OCT-2003 (TrEMBLrel. 25, Created)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Zinc finger protein-like protein.		
GN	P045303.32.		
OS	Oryza sativa (japonica cultivar-group).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Brhartoidae; Oryzae; Oryza.		
OX	NCBI_TaxID=39947;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Nipponbare;		
RA	Sasaki T., Matsumoto T., Yamamoto K.;		
RT	"Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC		

RT clones: P0453G03.1;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF004276; BAC79830.1; -- EDB82B36082A61F0 CRC64;
 SQ SEQUENCE 633 AA; 67636 MW; EDB82B36082A61F0 CRC64;

Alignment Scores:

Pred. No.: 0.303 Length: 633
 Score: 94.50 Matches: 36
 Percent Similarity: 37.69% Conservative: 13
 Best Local Similarity: 27.69% Mismatches: 40
 Query Match: 11.25% Indels: 41
 DB: 10 Gaps: 6

US-10-057-510-1 (1-486) x Q7188 (1-633)

QY 483 ATGGGGANTTTCCCTTTCTTCCAGGCTTTCGNAACACAGGGGGNCTANGANCACAGTTT 424
 DB 398 IIEGlyValAspThrPhe---LeuCysLeuAlaIleGlyAlaProSerSerLeuPhe 416
 QY 423 CCAGAACAGAGGGCCAGGACCAACACAGGGAGGGGCTTCTTATGCAAGTTCCCTTCA 364
 DB 417 ProGln-----ThraAlaSerAspProCysSerPheAlaProProAla 432
 QY 363 TGCCATATGAGCCAGNCTGTGGCTCGAACCTTGACTGGGCAATGCTCGAGAGTCT 304
 DB 433 ProHisMetSerAlaThr-----AlaLeuLeuGlnLysAla 444
 QY 303 GGGGACNATCCAGGGGCTGCTATCTTCATCTCCAGCCCTGCGAGGCTT---CTGCCCTCC 247
 DB 445 AlaGluValGlyAlaSerGlnSerSerSerPheLeuLysGluPheGlyLeuAlaAla 464
 QY 246 TCCAGGGGCTGCTCAATCTCTTCACT----- 220
 DB 465 SerThrSerSerSerProSerLysLeuSerGlnGlyArgPheThrThrGlyAsnThr 484
 QY 219 -----CCAGGGCTCTCCACAGCTCCCTCCACCTTCCACCTTCCACAGG 175
 DB 485 ProThrThrSerHisProHisProHisProHisProHis----- 497
 QY 174 ANACAGCGTGGCCCAAGGAGGAGGA 145
 DB 498 -----ProHisProGlnGly 503

RESULT 4

Q99491 PRELIMINARY; PRT; 191 AA.
 AC Q99491;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE DAN15 protein (Fragment).
 GN DAN15.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97051922; PubMed=8896557;
 RA Imbert G., Saudou F., Yvert G., Devys D., Trottier Y., Garnier J.M.,
 RA Weber C., Mandel J.L., Cancele G., Abbas N., Duerr A., Didierjean O.,
 RA Stevanin G., Agid Y., Brice A.;
 RT "Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with
 RT high sensitivity to expanded CAG/glutamine repeats.";
 RL Nat. Genet. 14:285-291(1996).
 DR EMBL; Y08266; CAA69592.1; --
 FT NON_TER 1
 FT NON_TER 191
 SQ SEQUENCE 191 AA; 20196 MW; E0A628C019583310 CRC64;

Alignment Scores:

Pred. No.: 0.473 Length: 191
 Score: 92.00 Matches: 38

Percent Similarity: 41.18% Conservative: 18
 Best Local Similarity: 27.94% Mismatches: 44
 Query Match: 10.95% Indels: 36
 DB: 4 Gaps: 5

US-10-057-510-1 (1-486) x Q99491 (1-191)

QY 450 AAACACAGGGGGNCTANGANCAGTTTTCAGAACAGAGGGCCAGGACCAACACAGGAG 391
 DB 9 LysSerGlyGly-----SerGluAlaAlaLeuLysGlu 19
 QY 390 GGGGCTTCTTATGCAAGTTCCTTCCATATGAGCCAGNCTGTGGCTCGAAC 331
 DB 20 GlySerAlaAlaAlaLeuSerSerSerSerSerAlaAlaAla----- 37
 QY 330 CTTCAGTGGGCAATGCTCGAGAGTCTGGGGAGCNATCCAGGGGCTGCACTTTCATCTCA 271
 DB 38 -----AlaAlaSerSerSerSerSerSerSerGlyProGlySerAlaMetGluThr--- 53
 QY 270 GCCCTGCGTAGCCCTTCTGCCCTCTCCAGGGCTGCTCCATCTCTTCAGCTCCAGGCC 211
 DB 54 -----GlyLeuLeuProAsnHisLysLeuLysThrValGlyGluAlaProAlaAla 70
 QY 210 CCTCCACACGTC-----CATGCCCTTC 190
 DB 71 ProProHisGlnGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHis 90
 QY 189 CACCTTCCACATGGANACAGCGCTGCCCCCAAGCCAGGAGGAGAACACAGAGCTGGCA 130
 DB 91 HisLeu--HisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 107
 QY 129 AGAAGAACAGCAGTGTAGCAGGGGCAACAATGGAGAAAGGAG 86
 DB 107 IINGINGINGINGINGINGINGINGINGINGINGINGINGINGINGINGING 121

RESULT 5

Q917W7 PRELIMINARY; PRT; 258 AA.
 AC Q917W7;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Minor structural protein.
 GN ORF3.
 OS Norwalk-like virus NLV/White River/290/1994/US.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
 OC Norovirus.
 OX NCBI_TaxID=171855;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Hu/NLV/White River/290/1994/US;
 RC MEDLINE=97193806; PubMed=9041391;
 RX Ando T., Monroe S.S., Noel J.S., Glass R.I.;
 RA "A one-tube method of reverse transcription-PCR to efficiently amplify
 RA a 3-kilobase region from the RNA polymerase gene to the poly(A) tail
 RA of small round-structured viruses (Norwalk-like viruses).";
 RL J. Clin. Microbiol. 35:570-577(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Hu/NLV/White River/290/1994/US;
 RC MEDLINE=2026071; PubMed=10804147;
 RX Ando T., Noel J.S., Fankhauser R.L.;
 RA "Genetic classification of 'Norwalk-like viruses.'";
 RL J. Infect. Dis. 181:S336-S348(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Hu/NLV/White River/290/1994/US;
 RC Ando T., Seto Y., Noel J.S., Monroe S.S., Glass R.I., Fankhauser R.L.;
 RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF414423; AAL13023.1; --
 DR InterPro; IPR004278; RNA_capsid.
 DR Pfam; PF03035; RNA_capsid; 1.
 DR SEQUENCE 258 AA; 27501 MW; 8D7B3CF2F64E6EAB CRC64;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazie R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow R.M., Basu A.B., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolabakov S.,
 RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RA "Sequencing of *Drosophila melanogaster* genome.";
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RA "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Flybase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AE003632; AAF53097.2; -.
 DR HSP; P15822; IBB0.
 DR Flybase; FBGN0004579; salm.
 DR GO; GO:0016481; p:negative regulation of transcription; NAS.
 DR GO; GO:0007438; p:ontocyte development; IMP.
 DR GO; GO:0007467; p:photoreceptor differentiation (sensu Drosop. .); IMP.
 DR GO; GO:0045485; p:P:R8 differentiation; NAS.
 DR GO; GO:0007433; p:sensory organ development; IMP.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; Zf-C2H2; 7.
 DR SMART; SM00355; Znf_C2H2; 7.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
 KW Metal-binding; Nuclear protein; Zinc; zinc-finger.
 SQ SEQUENCE 1355 AA; 148979 MW; D78505A058A1E451 CRC64;
 Alignment Scores:
 Pred. No.: 0.802 Length: 1355
 Score: 91.00 Matches: 46
 Percent Similarity: 41.40% Conservative: 19
 Best Local Similarity: 29.30% Mismatches: 63
 Query Match: 10.83% Indels: 29
 DB: 5 Gaps: 7
 US-10-057-510-1 (1-486) x Q9VKG2 (1-1355)
 QY 441 GGCTGANGCAGCTTTTCACAGACAGAGCCAGGACCAACAGGAGGGGCTTCT 382
 DB 1029 GlyValValAsnThrAsnProValArgProArgSerSerAlaSerHisGlyHisSer 1048
 QY 381 TATGCAAGT-----TCCCTTCATGCCATATGACCCAGNCTNGTGCT 337
 DB 1049 ValGlySerThrSerAlaProThrSerProSerValHisAlaSerSerGlnValIleLys 1068
 QY 336 CGAACCTTGACTCGGCAATGCTCGAGAGTCTGGGAC-----NATCCAGGG 289
 DB 1069 ArgSerSerProAlaArgSerGluAlaSerGlnGlyAlaLeuAspLeuThrProArg 1088
 QY 288 GCTGCA---TCTTCATCCTCAGCCCTCGTAGCCTTCTGCCCTCTCCAGGCTCGTCC 232
 DB 1089 AlaAlaProThrSerSerSerSerSerSerProLeuProLysGluLysProValSer 1108
 QY 231 ATTCCTTCAGCTCCAGGC-----CTCCTCCA 205
 DB 1109 ProProSerLeuPro-ArgSerProSerGlySerSerHisAlaSerAlaAsnIleLeuTh 1128
 QY 204 CACGTCCATGCTTCCACC-----TTGCCACATGGANACAGCCGTGCCCC 160
 DB 1128 rSerProLeuProProThrValGlyIleAspCysLeuProProGlyLeuGlnHis---Hi 1147
 QY 159 CACAGCCAGGAGGACAGACAGCTGGCAGAGAAAGAACAGACGATTAGACAGGGGCAAC 100
 DB 1147 sLeuGlnGlnGlnHisGlnHisLeuMetGlnGlnGlnAlaValAlaAlaAlaAla 1167
 QY 99 AATGGAGAAAGAGGATATCTATTAGCTTGGGCAATTTTACACCCAGCAC 51
 DB 1167 aAlaGlnHisHis---HisHisGlnGlnMetAlaAlaLeuHisGlnHis 1182
 RESULT 8
 Q9U412
 ID Q9U412 PRELIMINARY; PRT; 3469 AA.
 AC Q9U412;
 DC 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE SANT domain protein SMRTER.
 GN SMR OR SMRTER OR CG4013.
 OS *Drosophila melanogaster*.
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

```
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95417957; PubMed=10488333;
RA Tsai C.-C., Kao H.-Y., Yao T.-P., McKeown M., Evans R.M.;
RT "SMRTER, a Drosophila nuclear receptor coregulator, reveals that Ecr-
Mol. Cell 4:175-186(1999).
RL Mol. Cell 4:175-186(1999).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR EMBL; AF175223; AAD5264.1; -.
DR FlyBase; FBgn004309; Smr.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR001005; Myb DNA binding.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00717; SANT; 1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; 1.
DR DNA-binding; Nuclear protein.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 3469 AA; 364115 MW; 6284E14C5C247CD9 CRC64;

Alignment Scores:
Pred. No.: 0.918 Length: 3469
Score: 91.00 Matches: 48
Percent Similarity: 37.04% Conservative: 22
Best Local Similarity: 25.40% Mismatches: 60
Query Match: 10.83% Indels: 60
DB: 5 Gaps: 8

US-10-057-510-1 (1-486) x Q9U412 (1-3469)
QY 479 GGATNTTCCCTTTCTTCCAGGCTTTTCNAAAACAGGGGNGTANGANCAGTTTCCAG 420
Db 2128 GlyGlySerSerSerSerGly- SerGlySerGlyGlyProSerSerSer 2143
QY 419 AACAGAGGGCCAGACACACACAGGG- 393
Db 2144 AsparGHISHisGlyProProProProThrMetSerMetIleValArgSerGly 2163
QY 392 -----AGGGGGCTTTTATGCAAGTTCCTTCATCCCATATGCAAGGCTGCTG 342
Db 2164 GlyMetTyrArgGlyAspThrValThrValPro-----SerLeuAlaAlaProSerSer 2181
QY 341 TGGCTCGAACCTTGACT-----GGCAATGCTCTGAGAGT 306
Db 2182 TyrLeuTyrProThrArgSerValIleSerIleGlyGlyValValProGlyVal 2201
QY 305 CTGGGGACNATCCAGGGCTGCATCTTCATCTCAGGCTGCGT-----AGCCTT 256
Db 2202 LeuProGlyValProGlyIleAlaLeuTyrLeuGlnProValProValProIle 2221
QY 255 CTGCCCTCTCCAGGGCTGCTTCATCTTCACCTCAGGCTCTCTCCA----- 205
Db 2222 SerIleSerGlyGlnGlyGlnLeuProProIleGlnProProAlaGlnPro 2241
QY 204 -----CAC 202
Db 2242 ProSerGlyArgGlyValAlaValProProIleSerProGlnGlnAlaHisHis 2261
QY 201 GTCCATGCTTCCACCTTGGCCATGAGNACAGCGCTGCCCAAGAGGAGGAGACA 142
Db 2262 LeuHis-ProSer-----HisGlyHisSer-----ProSerGlnGlnGlnGln 2276
QY 141 ACGAGCTGGCAAGAAACAGCAGTGTAGACGAGGGGCAACATGAGAGAAAGAGTATC 82
Db 2276 NGInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2296
QY 81 TATTATGAGCTGGCAATTTACAC 57
Db 2296 uValIleSerGlySerIleIleHis 2304
```

RESULT 9

```
Q69115 PRELIMINARY; PRT; 209 AA.
AC Q69115;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE BZLF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88306247; PubMed=2841800;
RA Jensen H.B., Miller G.;
RT "Polymorphisms of the region of the Epstein-Barr virus genome which
RT disrupts latency.";
RL Virology 185:549-564(1988).
DR EMBL; M20821; AAA65492.1; -.
SQ SEQUENCE 209 AA; 22234 MW; EDE3CE3C8E27CFE0 CRC64;
```

Alignment Scores:

```
Pred. No.: 0.694 Length: 209
Score: 90.50 Matches: 33
Percent Similarity: 35.90% Conservative: 9
Best Local Similarity: 28.21% Mismatches: 48
Query Match: 10.77% Indels: 27
DB: 12 Gaps: 5
```

US-10-057-510-1 (1-486) x Q69115 (1-209)

```
QY 422 CAGACAGAGGCCAGGACCAACAGGAGGGGGCTTCTATGCAAGTCCCCTTCAT 363
Db 108 GlnAlaGlyGlyGluAlaProGlnProGlyAspAsnSerThrValGln---ProAlaAla 126
QY 362 GCATATGAGGCCAAGNCTGTGGCTCGAACCCCTTGACTGGGCAATGCTCTGAGAGTCTG 303
Db 127 AlaVal-----ValPheAlaCysPro----- 133
QY 302 GGGACNATCCAGGGCTGCATCTT-----CATCTCAGCCCTGGTAGCCCTTC 255
Db 134 GlyAlaAsnGlnGlyGlnGlnLeuAlaAspIleGlyAlaProGlnProAlaProAlaAla 153
QY 254 TGCCCTCTCTCCAGGGCTGCTTCATCTTCAGCTCCAGCGCTCTCCACACGCTCCATG 195
Db 154 AlaProAlaArgArgThrArgLysProLeuGln----- 164
QY 194 CTTTCCACCTTCCACATGGANACAGCGCTGCCCCCAAGCAGGAGGGAACAGAAAGC 135
Db 165 ProGluSerValArgArgGlySerLysArgValAlaAlaAsnGlyGlyAlaGlnArgGly 184
QY 134 TGGCAAGAAAGAACAGCAGTGTAGACGAGGGCAACAATGGAGAAAGAGTA 84
Db 185 TrpGlnArgThrValGlyLeuLysGluGlyGlySerGlyArgTrpGlyLeu 201
```

RESULT 10

```
Q69123 PRELIMINARY; PRT; 209 AA.
AC Q69123;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE BZLF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88306247; PubMed=2841800;
RA Jensen H.B., Miller G.;
```

RT "Polymorphisms of the region of the Epstein-Barr virus genome which
 RT disrupts latency";
 RL Virology 165:549-564(1988).
 DR ENBL: M2080; AAA65495.1; --
 SQ SEQUENCE 209 AA; 22278 MW; 876D395F4278CF59 CRC64;

Alignment Scores:
 Pred. No.: 0.694 Length: 209
 Score: 90.50 Matches: 32
 Percent Similarity: 35.04% Conservatives: 9
 Best Local Similarity: 27.35% Mismatches: 49
 Query Match: 10.77% Indels: 27
 DB: 12 Gaps: 4

US-10-057-510-1 (1-486) x Q69123 (1-209)

QY 422 CAGACAGAGCCAGGACACACAGGAGGGGCTTCTATGCAAGTCCCTTCAT 363
 DB 108 GlnAlaGlyGlyGluAlaProGlnProGlyAspAsnSerThrValGlnProGluAla 127
 QY 362 GCCATATGAGCCCAAGNCCTNGTGGCTCGAACCCCTTGACTGGGCAATGCTCTGAGAGTCTG 303
 DB 128 ValValPheAla-----CysPro----- 133
 QY 302 GGGACNATCCAGGGCTGCATCTT-----CATCTCAGCCCTGCGTAGCCCTC 255
 DB 134 GlyAlaAsnGlnGlyGlnGlnLeuAlaAspIleGlyAlaProGlnProAlaProAla 153
 QY 254 TGCCTCTCTCCAGGGCTCGTCATCTCCTTCAGCTCCAGCGCTCTCCACAGCGCCATG 195
 DB 154 AlaProAlaArgArgThrArgLysProLeuGln----- 164
 QY 194 CTTCCACCTTCCACATGGANACGCTGTCGCCCCACAGCCAGGAGGACAAACGAGC 135
 DB 165 ProGluSerValArgArgLysSerLysArgValAlaAlaAsnGlyGlyAlaGlnArgGly 184
 QY 134 TGCAGAGAAAGAACACAGCTTAGAGCAGGGGCAACAATGGAGAAAGAGTA 84
 DB 185 TrpGlnArgThrValGlyLeuLysGluGlyGlySerGlyArgTrpGlyLeu 201

RESULT 11

Q8XZN8 PRELIMINARY; PRT; 647 AA.
 ID Q8XZN8
 AC Q8XZN8;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2003 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE GALA protein 5.
 GN RSC1357 OR RS04639.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OC NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler C., Choise M., Moisan A., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Molan A., Robert C., Saurin W., Schiex T.,
 RA Siguer P., Theault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 DR EMBL: AL646064; CND15059.1; --
 DR InterPro; IPR001810; F-box.
 DR InterPro; IPR001611; LRR.
 DR Pfam; PF00646; F-box; 1.
 DR Pfam; PF00560; LRR; 2.
 DR SMART; SM00256; FBOX; 1.
 DR PROSITE; PS0181; FBOX; 1.
 KW Complete proteome.

SQ SEQUENCE 647 AA; 68312 MW; 8880B499A26F0F25 CRC64;

Alignment Scores:
 Pred. No.: 1.05 Length: 647
 Score: 89.50 Matches: 42
 Percent Similarity: 38.71% Conservatives: 18
 Best Local Similarity: 27.10% Mismatches: 57
 Query Match: 10.65% Indels: 38
 DB: 16 Gaps: 5

US-10-057-510-1 (1-486) x Q8XZN8 (1-647)

QY 486 CCCATGGGANTTCCCTTTCTTTCAGGCTTTTCNAAAACACAGGGGNCCTGANGNCAGT 427
 DB 53 ProGluGlyGlyPheProHis-----ValProAspGln 63
 QY 426 TTTCAGACAGAGGGCCAGGACACAAACAGGAGGGGGCTTCTATGCAAGTCCCTC 367
 DB 64 AlaGlyGluGlnArgAlaTyrProSerArgSerArgSerAlaGlyValSerSerSerPro 83
 QY 366 TCATGCCATATGAGCCCAAGNCCTNGTGGCTCGAACCCCTTGACTGGGCAATGCTCTGAGAG 307
 DB 84 -----LeuGlyGlyLeuAlaAlaMetArgLeuAsp----- 93
 QY 306 TCTGGGACNATCCAGGGCTGCATCTTCTCCTCAGCCCTGCTAGCCTTCTGCCCTCC 247
 DB 94 -----SerGlySerAlaGlyThrAlaProMetLeuArgSerArgLeuProPro 109
 QY 246 TCCAGGGCTCGTCCATCTCCTCAGCTCCAGCGCTCTCCACACGTCCTCATGCTTC-- 190
 DB 110 Ala-----ProSerAlaProSerGlyValProGlnValArgAsnIleGlu 124
 QY 189 -----CAGCTTGCACATGGANACAGCGCTGCCCCACACAG 154
 DB 125 GluMetProAlaGlyValLeuGlnHisValAlaSerPheLeuAspProArgSerArgArg 144
 QY 153 CCAGAGGGAACACAGAGCTGGCAGAAAGAACAGACAGCTTAGAGCAGGGGCAACAATGGA 94
 DB 145 -AlaLeuSerGlnValSerThrThrMetAsnAspAlaAlaArgSerSerGlnThrHisMe 164

RESULT 12

Q9HCL7 PRELIMINARY; PRT; 2414 AA.
 ID Q9HCL7
 AC Q9HCL7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein KIAA1555 (Fragment).
 GN KIAA1555.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20450693; PubMed=10997877;
 RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes.
 RT XVIII. The complete sequences of 100 new cDNA clones from brain which
 RT code for large proteins in vitro";
 RL DNA Res. 7:273-281(2000).
 DR EMBL: AB046775; BAB13381.2; --
 DR GO; GO:0016829; F-lyase activity; IEA.
 DR GO; GO:0008152; P-metabolism; IEA.
 DR InterPro; IPR002034; AIPM/Hcit_synth.
 DR InterPro; IPR007087; Znf.C2H2.
 DR Pfam; PF00096; zf-C2H2; 2.
 DR SMART; SM00355; Znf_C2H2; 5.

[illegible]

RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berhan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Butsum D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
 RA Fesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusekhin D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
 RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-F., Wasserman D.A., Weinstock G.M., Weissbach J., Wu D.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.M., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jatalui M., Kruse D., Li F., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Patel S., Pfeiffer B.,
 RA Pacleb J.J., Paragas V., Park S., Patel S., Richards S., Scheeler F.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnick S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berhan B., Carlson J.W., Ceiniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Asburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AEO3817; AAF58340.2; -;
 DR FlyBase; FBgn0033866; CG2680.
 DR SC SEQUENCE 1093 AA; 120895 MW; 3C722C76B363775F CRC64

Alignment Scores:

DR GO:0009231; P.vitamin B2 biosynthesis; IEA.
DR InterPro; IPR002606; FAD_Synth.
DR ProDom; PD003662; FAD_Synth; 1.
KW Hypothetical protein.
SQ SEQUENCE 1150 AA; 124288 MW; 208A27B908E7F870 CRC64;

Alignment Scores:
Pred. No.: 1.45 Length: 1150
Score: 88.50 Matches: 51
Percent Similarity: 35.88% Conservative: 10
Best Local Similarity: 30.00% Mismatches: 74
Query Match: 10.54% Indels: 36
DB: 3 Gaps: 8

US-10-057-510-1 (1-486) x Q872W5 (1-1150)

Qy 458 TTTCNAAAACACAGGGGNC-----TANGANCAGTTTTCACG 420
Db 407 PheGlySerGlyGlnGlySerAlaThrPheGlyAlaSerGlyAlaGlyProThrPheAsp 426
Qy 419 AACAGAGGG---CCAGGACCAACACAGGAGGGGGCTTCTTATGCAAGTTCCTTCAT 363
Db 427 GlyLeuGlyPheProGlyProProGly-----ProProProProArgAla 442
Qy 362 GCCATATGAGCCAGNCTNGTGGCTCGAACCC-----TTGACT 324
Db 443 AlaThrThrThrAlaAlaLeuProPheProProProProProProProProPro 462
Qy 323 GGGCATATGCTCAGAGTCTGGGACNATCCAGGGCTCATCTTCATCTCAGCCCTGC 264
Db 463 GlyProProGlnThrLeuGly---PheProGlyHisHisProProGlnLeuAla 481
Qy 263 GTAGCTTCTGCGCTCTCCAGGGCTCGTCCATTCCTTCAGCTCCAGCGCTCTCCAC 204
Db 482 GluAlaPheProProProProProProProProProProProProProProPro 501
Qy 203 ACCTCCATGCTTCCACCTTGGCCATATGGANACA-GCCTGCCCCA-----C 157
Db 501 isProHisLeuGlnHisAlaLysThrPheAlaLeuAlaLysProHisSerGlnGlnG 521
Qy 156 AAGCCAGAGGGAACAAGAGCTGGCAAGAACAGCAGCAGTATAGACAGGCGGCAACAAT 97
Db 521 lnProGlnLeuHisHisSerGlnSerLeuProGlnLeuGlnLeuGlnSerGlnSer 541
Qy 96 GGAGAAGGAGTATCTATTATGAGCTTG-----GCATTTTACA 58
Db 541 exArgLysGlnGluThrLysGlyPheThrSerAsnAlaLeuSerGluThrAlaTyLeu 561
Qy 57 CCAGC-----ACTTCCAGCTTTT 38
Db 561 laSerGlyLeuLeuSerHisProPhe 569

RESULT 15
Q82L04 PRELIMINARY; PRT; 378 AA.
ID Q82L04
AC Q82L04
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ATP/GTP-binding protein.
GN SAV2207.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VA-4680 / ATCC 31267 / NCIMB 12804 / NRRL B165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,
RT "Genome sequence of an industrial microorganism Streptomyces

Pred. No.: 1.44 Length: 1093
Score: 88.50 Matches: 40
Percent Similarity: 36.90% Conservative: 29
Best Local Similarity: 21.39% Mismatches: 70
Query Match: 10.54% Indels: 49
DB: 5 Gaps: 7

US-10-057-510-1 (1-486) x Q9V6T2 (1-1093)

Qy 485 CMTGGGGANTTTC-----CTTTCTCAGCTTTCNAAA 450
Db 899 ProTrpThrHisSerValAlaHisIleProProLysValLeuGlnAsnTyrThrArg 918
Qy 449 AACAGGGGNCNTANGANCAGTTTCCAGACAGAGGCGCAGGCCACACAGGGAGG 390
Db 919 MetLeuLysLysLeuProGlnLeuThrAlaLysThrProIleProGlnSerProLys 938
Qy 389 GGGCTTCTTATGCAAGTTCCTTCATGTCATATGAGCCAGNCTNGTGGCTCGAACCC 330
Db 939 AlaValLysLeuSerLysProValAlaAlaIle-----LysGlnProValLysPro 955
Qy 329 TTGACTGGGCAATGCTCTGAG-----AGTCTGGGACNATCCAGGGGCTGCAT 282
Db 956 ValArgGlyGluAlaProLysValAlaThrPheSerLeuAlaAspMetLys---PheAsp 974
Qy 281 CTTTCATCTCAGCCCTCGTAGCTTCTGCTCTCCAGGGCTCGTCCATTCCTTCAG 222
Db 975 LeuArgProPro-----ProProProGlnGlySerProTrpLeuGln 988
Qy 221 CTCAGGGCTCTCCACAGTCATGCTTCACCTTCCACATGAGANACGCGTGCC 162
Db 989 GlySerArgTyrIleTyrAsnThrLeuProAlaAspLeuAlaThrAlaAlaAlaSerThr 1008
Qy 161 CCCACAGGCCAGAG----- 148
Db 1009 ProThrLysMetValLysSerTyrIleGlyValValProAlaProThrSerAlaAsnSer 1028
Qy 147 -----GGAAACAGAGCTGGCAAGAAAGAACAGCAGTTAGACAGGGGCAACAATGGA 94
Db 1029 AspValGlyAspSerProLeuThrSerGluPheAspGluPheGlnArgTyrArgAsnGly 1048
Qy 93 GAA-----AGAGTATCTATTATGAGCTTGGCATTTT 61
Db 1049 HisSerLeuLysArgSerProAspPheGlyArgAsnLeuGlnLeuGluTyrGlyPheLys 1068
Qy 60 ACACACAGCATTCCAGCCTT 40
Db 1069 ProProLeuValProSerMet 1075

RESULT 14
Q872W5 PRELIMINARY; PRT; 1150 AA.
ID Q872W5
AC Q872W5
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein B2G14.030.
GN B2G14.030.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC Schulte U., Align V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Newes H.W., Mannhaupt G.,
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RL German Neurospora genome project;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX284753; CAD70448.1;
DR GO:0008531; F:riboflavin kinase activity; IEA

RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL: AP005029; BAC69919.1; -;
 DR InterPro: IPR005654; AFG1_ATPase.
 DR Pfam: PF03969; AFG1_ATPase; 1.
 KW Complete proteome.
 SQ SEQUENCE 378 AA; 40246 MW; BCD53E5ADAFE2356 CRC64;

Alignment Scores:

Pred. No.:	2.3	Length:	378
Score:	86.00	Matches:	37
Percent Similarity:	35.33%	Conservative:	16
Best Local Similarity:	24.67%	Mismatches:	37
Query Match:	10.24%	Indels:	60
DB:	16	Gaps:	7

US-10-057-510-1 (1-486) x Q82L04 (1-378)

QY	342	GTGGCTCGAACCTTGACTGGCAATGCTCTGAGAGTCTGGGACNATCCAGGGGCTGCA	283
DB	5	VaAlaAaArgInLeuGluTrp-----GlyThrVal	14
QY	282	TCITTCATCCCTCAGCCGTCGTAGCCTTCCTCCCTCCAGGGCTCGTCCATT-----	229
DB	15	SerSerSerThrAlaAlaSerGlyIleAspProIleAlaGluAlaProLeuSerLeu	34
QY	228	-----CCTTCAGCTCCAGCG-----CCTCTCCA	205
DB	35	CysAlaAArgAlaProHisValProAlaAspArgLeuValAlaGluMetValProPro	54
QY	204	CACGTCCATGCTCCACCTGTCACATGGANACAGCGGTGCCCCACACAGCCAGGGA	145
DB	55	ArgPheAspSerValArgPheAlaThrTyIle---ProAspProAsnGlnProSerGln	73
QY	144	ACA-----ACGAAGCTGGCAAGAAGAACAGCAGT	115
DB	74	ThrGluAlaValArgValLeuGluAspPheAlaThrGlyLeuGlyGlyAlaHisAlaSer	93
QY	114	TAGAGC-----AGGGCAACAAATGGGAA	91
DB	94	GlySerGlyArgArgLeuPheGlyPheGlyLysAlaProLysThrProAlaGlyPro	113
QY	90	AGGAGTATCTATTATGAG-----CTT	70
DB	114	ArgGlyValTyLeuAspGlyGlyTyGlyValGlyLysThrHisLeuLeuAlaSerLeu	133
QY	69	GGGCAATTTACACCGACCTCCAGCCTT	40
DB	134	TrpHisAlaThrProAlaGluProGluLeu	143

Search completed: April 6, 2004, 17:57:46
 Job time : 51.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 6, 2004, 17:48:49 ; Search time 14 Seconds
(without alignments)
3615.156 Million cell updates/sec

Title: US-10-057-510-1
Perfect score: 876
Sequence: 1 gcacgacgctatnaaaat.....aaagggaatccccatggg 486

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO.spool.p/US10057510/runat 06042004 141608 14709/app query.fasta_1.647
-DB=SwissProt 42 -OPMT=Eastan -SUFFIX=rsb -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10057510 @CGN_1_1_16 @runat 06042004 141608 14709 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	10.0	335	1 DUFF_MACMU	Q951f2 macaca mula
2	84.5	10.1	1355	1 SALM_DROME	P39770 drosophila
3	84	9.6	336	1 DUFF_HUMAN	Q16570 homo sapien
4	83.5	9.9	262	1 GDAL1_WHEAT	P04721 triticum ae
5	83.5	9.9	817	1 VRP1_YEAST	P37370 saccharomyc
6	83	9.9	213	1 STR5_HUMAN	Q9nsy2 homo sapien
7	83	9.5	336	1 DUFF_PANTR	Q951f3 pan troglod
8	83	9.9	761	1 TLE2_BRARE	O13166 brachydanio
9	83	9.5	2167	1 SKH1_RAT	Q9wv48 rattus norv
10	82.5	9.8	1161	1 KCH2_RABIT	Q8wny2 oryctolagus
11	82.5	9.8	1311	1 SPR4_CABEL	O17582 caenorhabdi
12	82	9.4	336	1 DUFF_PAPHA	Q951G5 papio hamad
13	82	9.8	503	1 CP3E_CAVPO	Q64417 cavia porce
14	82	9.8	503	1 CP3F_CAVPO	Q64406 cavia porce
15	82	9.8	1386	1 ZAP3_MOUSE	Q9r0i7 mus musculu
16	81	9.6	309	1 HX44_CHICK	P17277 gallus gall
17	81	9.2	336	1 DUFF_GORGO	Q951f9 gorilla gor
18	80.5	9.6	392	1 HME1_HUMAN	Q05925 homo sapien

C 19	80	9.5	265	1 IHA_SHEEP	P38440 ovis aries
C 20	80	9.5	360	1 IHA_BOVIN	P07994 bos taurus
C 21	80	9.5	540	1 GAG_MLVHO	P21435 homuly muri
C 22	79.5	9.5	268	1 CDX1_MOUSE	P18111 mus musculu
C 23	79.5	9.5	401	1 HME1_MOUSE	P09065 mus musculu
C 24	79.5	9.5	450	1 V50K_BYDVP	P09516 barley yell
C 25	78.5	9.3	815	1 PYGO_DROME	Q9v9w8 drosophila
C 26	78.5	9.3	1026	1 STAU_DROME	P25159 drosophila
C 27	78	9.3	834	1 SHTA_DROME	P28285 drosophila
C 28	77.5	9.2	1446	1 IE18_PRVKA	P33479 pseudorabie
C 29	77	9.2	234	1 ZEE3_HUMAN	Q961u2 homo sapien
C 30	77	9.2	852	1 WS14_HUMAN	Q9np71 homo sapien
C 31	77	9.2	1060	1 ZFH1_DROME	P28166 drosophila
C 32	76.5	8.7	652	1 ACSA_ECO57	Q8x5t5 escherichia
C 33	76.5	8.7	652	1 ACSA_ECOL6	Q8fay8 escherichia
C 34	76.5	8.7	652	1 ACSA_ECOL1	P27550 escherichia
C 35	76.5	9.1	1729	1 DME_ARATH	Q81K56 arabidopsis
C 36	76	9.0	428	1 FXB2_MOUSE	Q64733 mus musculu
C 37	76	8.7	2161	1 SHK1_HUMAN	Q9y566 homo sapien
C 38	75.5	9.0	137	1 YNAF_BAGSU	O31794 bacillus su
C 39	75.5	9.0	202	1 LPRE_MYCTU	Q11065 mycobacteri
C 40	75.5	9.0	269	1 TIR3_YEAST	P40552 saccharomyc
C 41	75.5	9.0	307	1 FMT_XANCP	Q8p490 xanthomonas
C 42	75.5	9.0	344	1 YBIE_SCHPO	P87179 schizosacch
C 43	75.5	9.0	544	1 H15_DROME	Q94890 drosophila
C 44	75.5	9.0	809	1 CN4D_HUMAN	Q08499 homo sapien
C 45	75	8.9	857	1 PAX6_DROME	O18381 drosophila

ALIGNMENTS

RESULT 1
DUFF_MACMU STANDARD; PRT; 335 AA.
ID_Q951f2;
AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DI 28-FEB-2003 (Rel. 41, Last annotation update)
DE Duffy antigen/chemokine receptor.
GN FY OR DARC.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RA Tournamille C., Blancher A., Le Van Kim C., Gane P., Apoil P.A.,
RA Nakamoto W., Catron J.-P., Colin Y.;
RT "Sequences, evolution and ligand binding properties of mammalian Duffy
antigen/receptor for chemokines";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Non-specific receptor for many chemokines (By
similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family Duffy of G-protein coupled
receptors.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF311921; AAL09456.1; -
DR InterPro; IPR005384; Duffyantigen.
DR PRINTS; PR01559; DUFFVANTIGEN.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 62 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 63 83 1 (POTENTIAL).
FT DOMAIN 84 94 CYTOPLASMIC (POTENTIAL).


```
FT TRANSMEM 95 115 2 (POTENTIAL).
FT DOMAIN 116 125 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 126 146 3 (POTENTIAL).
FT DOMAIN 147 165 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 166 186 4 (POTENTIAL).
FT DOMAIN 187 206 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 207 227 5 (POTENTIAL).
FT DOMAIN 228 243 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 244 264 6 (POTENTIAL).
FT DOMAIN 265 286 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 287 307 7 (POTENTIAL).
FT DOMAIN 308 335 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 16 16 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 26 26 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 335 AA; 35377 MW; D0FFB02DBB9A17D7 CRC64;

Alignment Scores:
Pred. No.: 2.07 Length: 335
Score: 88.00 Matches: 56
Percent Similarity: 38.50% Conservatve: 16
Best Local Similarity: 29.95% Mismatches: 55
Query Match: 10.05% Indels: 60
DB: 1 Gaps: 12

US-10-057-510-1 (1-486) x DUFF_MACMU (1-335)

QY 42 GGCTGGGAAGTGGTGGTGAATGCTCCCAAGCTCATATAGATCTCCT---TTCTCCAT 98
Db 94 GlyTrpProValuu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerI 110
QY 99 TGTTCCTCCCTCTTAACCTGCTGTTCTTCT-----TGCCAGCT 137
Db 110 eValValProIleLeuAlaProGlyLeuAlaProGlyLeu-ThrLeuGlyLeuSerLe 130
QY 138 TCGTGTGCTCCTCGG-----CTGTGGGG-----GC 164
Db 130 uGlyTyrcysValTrpTrpGlySerAlaPheAlaGlnAlaLeuLeuLeuGlyCysHisAl 150
QY 165 ACAGCTGTNTCCCA-----TGTGGCAAGGTGGAAGGATGAGCTGTGGAGGAGCG 215
Db 150 aSerLeuGlyProIleLeuGlyAlaGlyGlnValProGlyLeu-ThrLeuGlyLeuSerV 170
QY 216 CTGAGCTGAAGG-----ATGACAGAGCTCGGAGGAGGAGGAGGAGGAGGAGCTAC 263
Db 170 alGlyLeuTrpGlyValAlaAlaLeuLeu-ThrLeuProIleThrLeuAlaSerGlyAla 189
QY 264 GCAGGG-----CTGAGGATGAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 305
Db 190 SerGlyGlyLeuGlyThrProAlaTyrcysMetGluLeuLeuAlaGlnAlaThrHis 209
QY 306 ACTCTCAGGACATGTCCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 362
Db 210 AlaValAlaCysLeuAlaValPheValLeuLeuProLeuGlyLeuPheGlyAlaLysGly 229
QY 363 ATGAAGGGAACTTGCATAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 413
Db 230 LeuLysLysAlaLeuGlyMetGlyProGlyProIleProIleMetAsnIleLeuTrpAlaTrp-Phe 249
QY 414 TCTGTTCTGGAAGAACTGGTTCNTAGNCCCTCGTTTTCGCAAGAGGAGGAGGAGGAGGAG 473
Db 249 eIlePheTrp-----Trp-----253
QY 474 AANTCCCATGGG 486
Db 254 -----ProHisGly 256

RESULT 2
ID_SALM_DROME STANDARD; PRT; 1355 AA.
AC P39770;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
```

```
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeotic protein spalt-major.
GN SALM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94139659; PubMed=7905822;
RA Kuehnlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
RA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaeckle H., Schuh R.;
RT "Spalt encodes an evolutionarily conserved zinc finger protein of
RT novel structure which provides homeotic gene function in the head and
RT tail region of the Drosophila embryo."
RL EMBO J. 13:168-179(1994).
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST
CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE
CC TRANSCRIPTION OF THE TSH GENE.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -!- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC use European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75541; CAA53229.1; -.
CC PIR; S40022; S40022.
CC DR HSPP; P15822; 1BBO.
CC DR FlyBase; FBgn0004579; salm.
CC DR GO; GO:0016481; P:negative regulation of transcription; NAS.
CC DR GO; GO:0007438; P:photocycle development; IMP.
CC DR GO; GO:0007487; P:photoreceptor differentiation (sensu Drosop. .); IMP.
CC DR GO; GO:0045465; P:R8 differentiation; NAS.
CC DR GO; GO:0007423; P:sensory organ development; IMP.
CC DR InterPro; IPR007087; Znf_C2H2.
CC DR Pfam; PF00096; zf-C2H2; 7.
CC DR SMART; SM00355; Znf_C2H2; 7.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
CC DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
CC DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
KW Developmental protein; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Transcription regulation; Repeat.
FT ZN_FING 451 473 C2H2-TYPE 1.
FT ZN_FING 479 501 C2H2-TYPE 2.
FT ZN_FING 824 846 C2H2-TYPE 3.
FT ZN_FING 852 874 C2H2-TYPE 4.
FT ZN_FING 884 906 C2H2-TYPE 5.
FT ZN_FING 1289 1311 C2H2-TYPE 6.
FT ZN_FING 1317 1339 C2H2-TYPE 7.
SQ SEQUENCE 1355 AA; 148995 MW; 6FC4EBDC0BC6355E CRC64;

Alignment Scores:
Pred. No.: 4.82 Length: 1355
Score: 84.50 Matches: 46
Percent Similarity: 39.29% Conservatve: 20
Best Local Similarity: 27.38% Mismatches: 56
Query Match: 10.06% Indels: 36
DB: 1 Gaps: 7

US-10-057-510-1 (1-486) x SALM_DROME (1-1355)
QY 441 GGNCTGANGNCAGTTTTCAGACAGAGGCCAGGCCACCAACAGGAGGGGCTTCT 382
```

```

Db      1029 GlyValValAsnThrAsnProValArgSerSerAlaSerSerHisGlyHisSer 1048
QY      381 TATGCAAGT-----TCCCTTCATCCATATGAGCCAAAGNCTGTGCT 337
Db      1049 ValGlySerThrSerAlaProThrSerProSerValHisAlaSerSerGlnValIleLys 1068
QY      336 CCAACCCCTGATGGGCATGCTCCAGAGTCTGGGGAC-----NATCCAGGG 289
Db      1069 ArgSerSerProAlaArgSerGluAlaSerGlnGlyAlaLeuAspLeuThrProArg 1088
QY      288 GTGCA--TCTTCATCTCCAGCCCTGCGTAGCCTTCTGCCCTCTCCAGGGCTCGTCC 232
Db      1089 AlaAlaProThrSerSerSerSerArgSerProLeuProLysGluLysProValSer 1108
QY      231 ATTCTTCAGTCCAGCCG-----TTGCCATGATGGANACAGCGGTGCCCC 160
Db      1109 ProProSerLeuProArgSerProSerGlySerHisAlaSerAlaAsnIleLeuTh 1128
QY      204 CACGTCCATGCTTCCACC-----TTGCCATGATGGANACAGCGGTGCCCC 160
Db      1128 rSerProLeuProProThrValGlyLeaspCysLeuProLysGlyLeuGlnHis---Hi 1147
QY      159 CACAGCCAGAGGGAACAGAGCTGGCAAGAAAGACAGCAGTAGTAGAGGGGGCAAC 100
Db      1147 sleuGlnGlnGlnHisGlnHisLeuMetGlnGlnGlnAlaValAlaAlaAlaAla 1167
QY      99 AATGGAGAAGAGGATATCTATTATGAGCTGGGCATTTTACACAGCAGCATTCCAGCCCTT 40
Db      1167 alaGlnHisHis-----HisHisGlnGlnMetAlaAlaLe 1179
QY      39 TTCCATAAATCGGAACATT 18
Db      1179 uaspGlnHisGlnGlnGlnLeu 1186

RESULT 3
DUFF_HUMAN
ID DUFF_HUMAN STANDARD; PRT; 336 AA.
AC Q16570; O75898; Q16300; Q8WME3; Q9UUF0; Q9UKZ5; Q9UKZ6; Q9UOE1;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Duffy antigen/chemokine receptor (Fy glycoprotein) (gpFy)
DE (Glycoprotein D) (Plasmodium vivax receptor) (CD234 antigen).
DE FY OR DARC OR GDP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.
RP TISSUE=Blood;
RX MEDLINE=94068488; PubMed=8248172;
RA Chaudhuri A., Polyakova J., Zbrzezna V., Williams K., Gulati S.,
RA Pogo A.;
RT "Cloning of glycoprotein D cDNA, which encodes the major subunit of
RT the Duffy blood group system and the receptor for the Plasmodium
RT vivax malaria parasite.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10793-10797(1993).
RN [2]
SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Blood;
RX MEDLINE=95392577; PubMed=7663520;
RA Tournamille C., Colin Y., Carttron J.-P., Le van Kim C.;
RT "Disruption of a GATA motif in the Duffy gene promoter abolishes
RT erythroid gene expression in Duffy-negative individuals.";
RL Nat. Genet. 10:224-228(1995).
RN [3]
SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT GLX-42.
RP TISSUE=Blood;
RX MEDLINE=95134891; PubMed=7833467;
RA Iwamoto S., Omi T., Kajii E., Ikemoto S.;
RT "Genomic organization of the glycoprotein D gene: Duffy blood group

```

```

RT Fya/Fyb alloantigen system is associated with a polymorphism at the
RT 44-amino acid residue.";
RL Blood 85:622-626(1995).
RN [4]
SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS CYS-89 AND THR-100.
RP TISSUE=Blood;
RX MEDLINE=98402363; PubMed=9731074;
RA Tournamille C., Le Van Kim C., Gane P., Le Pennec P.Y., Roubinet F.,
RA Babinet J., Carttron J.-P., Colin Y.;
RT "Arg89Cys substitution results in very low membrane expression of the
RT Duffy antigen/receptor for chemokines in Fy(x) individuals.";
RL Blood 92:2147-2156(1998).
RN [5]
SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS CYS-89 AND THR-100.
RP MEDLINE=99101431; PubMed=9886340;
RA Olsson M.L., Smythe J.S., Hansson C., Poole J., Mallinson G.,
RA Jones J., Arent J.D., Daniels G.;
RT "The Fy(x) phenotype is associated with a missense mutation in the
RT Fy(b) allele predicting Arg89Cys in the Duffy glycoprotein.";
RL Br. J. Haematol. 103:1184-1191(1998).
RN [6]
SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLX-42.
RP Begguiley C.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLY-42.
RP TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [8]
SEQUENCE OF 1-117 FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT GLY-42.
RX MEDLINE=20040661; PubMed=10570183;
RA Zimmerman P.A., Woolley I., Masinde G.B., Miller S.M., McNamara D.T.,
RA Hazlett F., Mgone C.S., Alpers M.P., Genton B., Boatín B.A.,
RA Kazura J.W.;
RT "Emergence of FYA*(null) in a Plasmodium vivax-endemic region of
RT Papua New Guinea.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13973-13977(1999).
RN [9]
VARIANT GLX-42.
RP TISSUE=Periphereal blood;
RX MEDLINE=95220914; PubMed=7705836;
RA Tournamille C., Le van Kim C., Gane P., Carttron J.-P., Colin Y.;
RT "Molecular basis and PCR-DNA typing of the Fya/fyb blood group
RT polymorphism.";
RL Hum. Genet. 95:407-410(1995).
RN [10]
VARIANT CYS-89.
RX MEDLINE=98421365; PubMed=9746760;
RA Parasol N., Reid M., Rios M., Castilho L., Harari I., Kosower N.S.;
RT "A novel mutation in the coding sequence of the Fy* allele of the
RT Duffy chemokine receptor gene is associated with an altered
RT erythrocyte phenotype.";
RL Blood 92:2237-2243(1998).

```

CC -!- FUNCTION: Non-specific receptor for many chemokines such as IL-8,
 CC GRO, RANTES, MCP-1 and TARC. It is also the receptor for the human
 CC malaria parasites Plasmodium vivax and Plasmodium knowlesi.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=2;
 CC IsoId=Q16570-1; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q16570-2; Sequence=VSP_001323;
 CC -!- TISSUE SPECIFICITY: Found in adult kidney, adult spleen, bone
 CC marrow and fetal liver. In particular, it is expressed along
 CC postcapillary venules throughout the body, except in the adult
 CC liver. Erythroid cells and postcapillary venule endothelium are
 CC the principle tissues expressing Duffy. Fy(-A-B) individuals do
 CC not express Duffy in the bone marrow, however they do, in
 CC postcapillary venule endothelium.
 CC -!- POLYMORPHISM: DY is responsible for the Duffy blood group system.
 CC The molecular basis of the Fy(A)=Fy1/Fy(B)=Fy2 blood group
 CC antigens is a single variation in position 42; Gly-42 corresponds
 CC to Fy(A) and Asp-42 to Fy(B).
 CC -!- DISEASE: Individuals that do not produce the Duffy antigen
 CC (Fy(A-B-)) are more resistant to vivax malaria. This allele is
 CC found predominantly in population of african origin.
 CC -!- SIMILARITY: Belongs to family Duffy of G-protein coupled
 CC receptors.
 CC -!- DATABASE: NAME=PROW; NOTE=PROW 2:90-94(2001);
 CC WWW='http://www.ncbi.nlm.nih.gov/prow/guide/303294872.g.htm'.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; U01839; AAC50055.1; -;
 CC DR EMBL; X85785; CAB59770.1; -;
 CC DR EMBL; S76830; CAB33239.1; -;
 CC DR EMBL; AF055992; AAC72301.1; -;
 CC DR EMBL; AF030521; AAD20435.1; -;
 CC DR EMBL; AL035403; CAB56228.1; -;
 CC DR EMBL; BC017817; AAF17817.1; -;
 CC DR EMBL; AF100634; AAF02415.1; -;
 CC DR EMBL; AF100634; AAF02416.1; -;
 CC DR PIR; I52608; I52608.
 CC DR Genew; HGNC:4035; FY.
 CC DR MIM; 110700; -;
 CC DR GO; GO:0005886; C:plasma membrane; TAS.
 CC DR GO; GO:0004888; F:transmembrane receptor activity; TAS.
 CC DR InterPro; IPR005384; Duffyantigen.
 CC DR PRINTS; PR01559; DUFFYANTIGEN.
 CC KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 CC Alternative splicing; Polymorphism; Blood group antigen.
 CC FT DOMAIN 1 63
 CC FT EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 64 84
 CC FT 1 (POTENTIAL).
 CC FT DOMAIN 85 95
 CC FT CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 96 116
 CC FT 2 (POTENTIAL).
 CC FT DOMAIN 117 126
 CC FT EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 127 147
 CC FT 3 (POTENTIAL).
 CC FT DOMAIN 148 166
 CC FT CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 167 187
 CC FT 4 (POTENTIAL).
 CC FT DOMAIN 188 207
 CC FT EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 208 228
 CC FT 5 (POTENTIAL).
 CC FT DOMAIN 229 244
 CC FT CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 245 265
 CC FT 6 (POTENTIAL).
 CC FT DOMAIN 266 287
 CC FT EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 288 308
 CC FT 7 (POTENTIAL).
 CC FT DOMAIN 309 336
 CC FT CYTOPLASMIC (POTENTIAL).
 CC FT CARBOHYD 16 16
 CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 27 27
 CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 33 33
 CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPLIC 1 7 MGNCLHR -> MASSGVVLQ (in isoform 1).
 FT FT FTID=VSP_001323.
 FT VARIANT 42 42 D -> G (in Fy(a) antigen; dbSNP:12075).
 FT FTID=VAR_003480.
 FT VARIANT 89 89 R -> C (in Fy(x)).
 FT FTID=VAR_015068.
 FT VARIANT 100 100 A -> T (in dbSNP:13962).
 FT FTID=VAR_015069.
 FT CONFLICT 326 326 S -> F (in REF. 7).
 SQ SEQUENCE 336 AA; 35610 MW; ACCEA15FBFICBE9B CRC64;
 Alignment Scores:
 Pred. No.: 4.82 Length: 336
 Scores: 84.00 Matches: 54
 Percent Similarity: 37.23% Conservative: 16
 Best Local Similarity: 28.72% Mismatches: 56
 Query Match: 9.59% Indels: 62
 DB: 1 Gaps: 11
 US-10-057-510-1 (1-486) x DUFF_HUMAN (1-336)
 QY 42 GCGTGGGAAGTGTGGTGAATAATGCCCAAGCTCATATAGATACCTCT---TTCTCCAT 98
 Db 95 GlyTrpProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerIl 111
 QY 99 TGTTCGCCCTGCTCTAACTGCTGTCTCTCTCT-----TGCCAGCT 137
 Db 111 eValValProValLeuAlaProGlyLeuGlySerThrArgSerSerAlaLeuCysSerLe 131
 QY 138 TCGTTGTTCCTCTGGCTT----- 156
 Db 131 uGlyTyxCysValTrpTyrglySerAlaPheAlaGlnAlaLeuLeuGlyCysHisAl 151
 QY 157 ---GTGGGSCACGCTGTNTCCATGTGCAAGTGGCAAGCATGACGCTGTGGAGGAG 212
 Db 151 aSerLeuGlyHisArgLeuGly---AlaGlyGlnValProGlyLeu-ThrLeuGlyLeu 170
 QY 213 GCGTGGAGCTGAAGGA-----ATGGACAGACGCTGGGAGGAGGGCAGAGGC 260
 Db 170 hrValGlyIleTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGly 189
 QY 261 TACGACGG-----CTGAGGATGAAGATGCAGCCCTGGATNGTCCC 302
 Db 190 AlaSerGlyGlyLeuCysThrLeuIleTy:SerThrGluLeuTySAlaLeuGlnAlaThr 209
 QY 303 CAGACTCTCAGACATTGCCAGTCAAGGGTTCGAGCCACNAGNCNTT---GGCTCATAT 359
 Db 210 HisThrValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLys 229
 QY 360 GGCATGAAGGGAACCTTGCATAAGAGCCCTCCCTGG-----TTGTGCTCCTGG 410
 Db 230 GlyLeuLysLysAlaLeuGlyMetGlyProGlyProTrpMetAlaIleLeuTrpAlaTrp 249
 QY 411 CCCTCTGTCTCGAAAACTGGTCTAGNCCCTCGCTTTTNGCAAAGCCTGAAGAAG 470
 Db 250 -PheIlePheTrp-----Tip----- 254
 QY 471 GGAAATCCCATCGG 486
 Db 255 -----ProHisGly 257
 RESULT 4
 GDAL WHEAT
 ID GDAL WHEAT STANDARD; PRT; 262 AA.
 AC P04721;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, last sequence update)
 DT 15-JUL-1999 (Rel. 38, last annotation update)
 DE Alpha/beta-gliadin A-1 precursor (Prolamin).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticaceae; Triticum.


```
FT DOMAIN 704 708 POLY-PRO.
FT CONFLICT 308 308 P -> R (IN REF. 1).
FT CONFLICT 350 350 A -> R (IN REF. 1).
FT CONFLICT 689 689 V -> E (IN REF. 1).
FT CONFLICT 710 817 PSTMDTGTSTNSPSKLNKORLFTSGSTLOKHRTHTNQPDV
FT DVERVYTGGSNSIVGAKSGNERLVIDDSFKWTNVSOMKPK
FT RPFONTKLPSPGKGSVPLDLTFT -> HUNIPVPLIA
FT FVKTLNNGYFLQVDRRCNTSIIRIQINQMLM (IN REF.
FT 1).
SQ SEQUENCE 817 AA; 82593 MW; 24C7522D5B1CA1C8 CRC64;

Alignment Scores:
Pred. No.: 5.73 Length: 817
Score: 83.50 Matches: 23
Percent Similarity: 48.57% Conservative: 11
Best Local Similarity: 32.86% Mismatches: 15
Query Match: 9.94% Indels: 21
DB: Gaps: 2

US-10-057-510-1 (1-486) x VRP1_YEAST (1-817)
Qy 294 CCAAGGCTGCATCTTCATCTCAGCCCTCGAGCCTCTG----- 253
Db 406 ProglyAlaPheSerThrSerAlaLeuSerAlaSerIleProLeuAlaProLeu 425
Qy 252 -----CCCTCTCCAGGCTCCTCATCTTCAGTCAGCGCTCTCCA 205
Db 426 ProProProProProSerValAlaThrSerValProSerAlaProProProPro 445
Qy 204 CACGTCATGCTTCCACCTTGCACATGGANAGCCGCTGCCCCACAGCAGAGGGA 145
Db 446 ThrLeu-----ThrThrAsnLysProSerAla 454
Qy 144 ACACGAGCTGGCAAGAAAGACAGCAGT 115
Db 455 SerSerLysGlnSerLysIleSerSer 464

RESULT 6
ID STR5_HUMAN STANDARD; PRT; 213 AA.
AC Q9NSY2; P59094;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE STAR-related lipid transfer protein 5 (STARD5) (START domain-
DE containing protein 5).
GN STARD5
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Liver;
EX MEDLINE=22008095; PubMed=12011452;
RA Soccio R.E., Adams R.M., Romanowski M.J., Sehayek E., Burley S.K.,
RA Breslow J.L.;
RT "The cholesterol-regulated stard4 gene encodes a STAR-related lipid
RT transfer protein with two closely related homologues, Stard5 and
RT StarD6."
RL Proc. Natl. Acad. Sci. U.S.A. 99:6943-6948 (2002).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Testis;
RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Small intestine;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
```

```
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Mullahy S.J.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Gunaratne P.H.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R.W., Touchman J.W., Green E.D., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: May be involved in the intracellular transport of
CC sterols or other lipids. May bind cholesterol or other sterols (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC Name=2;
CC IsoId=Q9NSY2-1; Sequence=Displayed;
CC IsoId=Q9NSY2-2; Sequence=VSP_006278;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Contains 1 START domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF480304; AAL89654.1; -
DR EMBL; AL137657; CAB70862.1; -
DR EMBL; AK026352; -; NOT ANNOTATED CDS.
DR EMBL; BC004365; AAH04365.1; ALT_INIT.
DR FIC; T46357; T46357 STARD5.
DR Genew; HGNC:18065; STARD5.
DR MIM; 607050; -
DR InterPro; IPR000799; STAR.
DR InterPro; IPR002913; STAR.
DR Pfam; PF01852; START; 1.
DR PRINTS; PR00978; STARPTEIN.
DR SMART; SM02234; START; 1.
DR PROSITE; PS00448; START; 1.
KW Transpos; Lipid transport; Lipid-binding; Alternative splicing.
FT DOMAIN 1 213
FT VARSPLIC 1 107 Missing (in isoform 2).
FT SEQUENCE 213 AA; 23794 MW; EE5547A7846AB3C4 CRC64;

Alignment Scores:
Pred. No.: 5.75 Length: 213
Score: 83.00 Matches: 40
Percent Similarity: 33.55% Conservative: 12
Best Local Similarity: 25.81% Mismatches: 60
Query Match: 9.88% Indels: 43
DB: Gaps: 6

US-10-057-510-1 (1-486) x STR5_HUMAN (1-213)
```

QY 456 TGCAAAACCA-----GGGNGCTGANGCAGTTTTCAGACAG----- 415
 Db CysValLysProAlaValGlyLeuArgValLysTrpAspGluAsnValThrGlyPhe 86
 QY 414 -----AGGCCAGCACCAACACAGGAGGG 388
 Db GluIleIleGlnSerIleThrAspThrLeuCysValSerArgThrProSerAla 106
 QY 387 GCTTCTTATGCAAGTTCCTTCATGCCATATGAGCCAGNCCTGCTCGAACCCCT 328
 Db AlaMetLysLeuIleSerProArgAspPheValAspLeuValLysArgTyrGlu 126
 QY 327 GACTGGGCA-----TGTCTGAGAGTCT 304
 Db Asp-GlyThrIleSerSerAsnAlaThrHisValGluHisProLeuCysProProLysPr 146
 QY 303 GGGGACNATCCAGGGCTCATCTTCATCTCAGCCCTGCGTAGCTTCTGCGCTCC 244
 Db OglyPheValArgGlyPheAsn-----HisProCysGlyCysPheCysGluProLe 163
 QY 243 CAGGGCTGCTCATTC-----CTTCAGCTCAGCGCTCTCCACAGCTCCATCCCTCCAC 187
 Db uProGlyGluProThrLysThrAsnLeuValThrPhePheHisThrAspLeuSerGlyTy 183
 QY 186 CTTGCCACATGGANACAGCGCTGCCCCACAGCCAGGAGGAGGACACGAAGCTGCCAGA 127
 Db rLeuPro-----GlnAsnValValAspSerPhePr 194
 QY 126 AAGACAGCAGTGTAGCAGCGGCAACAATGAGAAAGGAGTA 84
 Db aArgSerMetThrArgPheTyrAlaAsnLeuGlnLysAlaVal 208

RESULT 7
 DUFF_PANTR STANDARD; PRT; 336 AA.
 AC Q95LF3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Duffy antigen/chemokine receptor.
 GN FY OR DARC.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tournamille C., Blancher A., Le Van Kim C., Gane P., Apoll P.A.,
 RA Nakamoto W., Cartron J.-P., Colin Y.,
 RT "Sequences, evolution and ligand binding properties of mammalian Duffy
 RT antigen/receptor for chemokines."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Non-specific receptor for many chemokines (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family Duffy of G-protein coupled
 CC receptors.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF311920; AAL09458.1;
 CC InterPro; IPR005384; Duffyantigen.
 CC PRINTS; PRO1559; DUFFYANTIGEN.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 63 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 64 84
 FT TRANSMEM 64 84 1 (POTENTIAL).

DOMAIN 85 95 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 96 116 2 (POTENTIAL).
 DOMAIN 117 126 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 127 147 3 (POTENTIAL).
 DOMAIN 148 166 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 167 187 4 (POTENTIAL).
 DOMAIN 188 207 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 208 228 5 (POTENTIAL).
 DOMAIN 229 244 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 245 265 6 (POTENTIAL).
 DOMAIN 266 287 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 288 308 7 (POTENTIAL).
 DOMAIN 309 336 CYTOPLASMIC (POTENTIAL).
 CARBOHYD 16 16 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 336 AA; 35625 MW; 23067E1D8BF7CBF2 CRC64;

Alignment Scores:
 Pred. No.: 5.95 Length: 336
 Score: 83.00 Matches: 54
 Percent Similarity: 37.23% Conservative: 16
 Best Local Similarity: 28.72% Mismatches: 56
 Query Match: 9.47% Indels: 62
 DB: 1 Gaps: 11

US-10-057-510-1 (1-486) x DUFF_PANTR (1-336)

QY 42 GGCTGGGAAGTGTGGTGTAAATGCCAAGCTCATATAGATACACTCT---TTCTCCAT 98
 Db 95 GlyTrpProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerIle 111
 QY 99 TGTGGCCCTGCTCAACTGCTGCTCTTCT-----TGCAGCT 137
 Db 111 eValValProIleLeuAlaProGlyLeuGlySerThrArgSerAlaLeuCysSerLe 131
 QY 138 TCCTTGTCTCTCTGCTT----- 156
 Db 131 uGlyTyrCysValTyrTyrGlySerAlaPheAlaGlnAlaLeuLeuGlyCysHisAl 151
 QY 157 -----GTGGGGCACGGCTGNTCCATGTGGCAAGGTGGAGGATGGAGTGTGGAGGAG 212
 Db 151 aSerLeuGlyHisArgLeuGly---AlaGlyGlnValProGlyLeu-ThrLeuGlyLeuT 170
 QY 213 GCCTCGAGCTGAAGGA-----ATGCACGAGCCCTGGGAGGAGGGGAGAGGC 260
 Db 170 hrValGlyIleTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGly 189
 QY 261 TAGCCAGGG-----CTGAGATGAGATGAGCCCTCGATNGTCC 302
 Db 190 AlaSerGlyGlyLeuCysThrLeuIleTyrSerThrGluLeuLysAlaLeuGlnAlaThr 209
 QY 303 CAGACTCTCAGGACATTGCCAGTCAAGGTTTCGAGCCACNAGNCTT---GGCTCATAT 359
 Db 210 HisThrValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLys 229
 QY 360 GCATGAAGGGAACTTGCATAGAAGCCCTCCCTGG-----TTGTGCTCTGG 410
 Db 230 GlyLeuLysAlaLeuGlyMetGlyProGlyProTyrMetSerIleLeuTrpAlaTrp 249
 QY 411 CCTCTGTTCGGAAACTGGTCTAGNCCCTCCCTGGTTTTCGAAAGCCCTGAGAAAG 470
 Db 250 -PheIlePheTrp-----Tyr----- 254
 QY 471 GGAANTCCCATGGG 486
 Db 255 -----ProHisGly 257

RESULT 8
 TLE2_BRARE
 ID TLE2_BRARE STANDARD; PRT; 761 AA.
 AC Q13166;
 DT 15-JUL-1998 (Rel. 36, Created)

RP DEVELOPMENTAL STAGE.
 RC TISSUE=Brain;
 RX MEDLINE=99436166; PubMed=10506216;
 RA Lim S., Naisbitt S., Yoon J., Hwang J.-I., Suh P.G., Sheng M., Kim E.;
 RT "Characterization of the Shank family of synaptic proteins. Multiple
 RT genes, alternative splicing, and differential expression in brain and
 RT development." J. Biol. Chem. 274:29510-29518(1999).
 RL J. Biol. Chem. 274:29510-29518(1999).
 RN [5]
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Brain;
 RX MEDLINE=20020275; PubMed=10551867;
 RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
 RT "Sonata protein receptor interacting protein defines a novel family of
 RT multidomain proteins present in human and rodent brain." J.
 RL J. Biol. Chem. 274:32997-33001(1999).
 RN [6]
 RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
 RX MEDLINE=99360651; PubMed=10433269;
 RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,
 RA Doan A., Asakura V.K., Lanahan A.A., Sheng M., Worley P.F.;
 RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
 RT postsynaptic density proteins." J. Biol. Chem. 276:40104-40112(2001).
 RL J. Biol. Chem. 276:40104-40112(2001).
 RN [8]
 RP INTERACTION WITH SPTAN1.
 RX MEDLINE=21523912; PubMed=11509555;
 RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,
 RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;
 RT "Synaptic scaffolding proteins in rat brain. Ankryrin repeats of the
 RT multidomain Shank protein family interact with the cytoskeletal
 RT protein alpha-fodrin." J. Biol. Chem. 276:40104-40112(2001).
 RL J. Biol. Chem. 276:40104-40112(2001).
 RN [9]
 RP FUNCTION.
 RX MEDLINE=21389514; PubMed=11498055;
 RA Sala C., Piech V., Wilson N.R., Passafaro M., Liu G., Sheng M.;
 RT "Regulation of dendritic spine morphology and synaptic function by
 RT Shank and Homer." J. Neurosci. 21:115-130(2001).
 RL J. Neurosci. 21:115-130(2001).
 RN [9]
 RP REVIEW.
 RX MEDLINE=20267867; PubMed=10806096;
 RA Sheng M., Kim E.;
 RT "The Shank family of scaffold proteins." J. Cell Sci. 113:1851-1856(2000).
 RL J. Cell Sci. 113:1851-1856(2000).
 CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic
 CC density (PSD) of excitatory synapses that interconnects receptors
 CC of the postsynaptic membrane including NMDA-type and metabotropic
 CC glutamate receptors, and the actin-based cytoskeleton. May play a
 CC role in the structural and functional organization of the
 CC dendritic spine and synaptic junction. Overexpression promotes
 CC maturation of dendritic spines and the enlargement of spine heads
 CC via its ability to recruit Homer to postsynaptic sites, and
 CC enhances presynaptic function.
 CC -!- SUBUNIT: May homomultimerize via its SAM domain. Interacts with
 CC SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with
 CC DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via
 CC the PDZ domain (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
 CC neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9WV48-1; Sequence=VSP_006072, VSP_006073;
 CC Name=2;
 CC IsoId=Q9WV48-2; Sequence=VSP_006072, VSP_006073;
 CC Name=3;
 CC IsoId=Q9WV48-3; Sequence=VSP_006074;
 CC Name=4; Synonyms=A;
 CC IsoId=Q9WV48-4; Sequence=VSP_006075;
 CC Name=5;

CC IsoId=Q9WV48-5; Sequence=VSP_006076, VSP_006077;
 CC TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,
 CC CA1 region hippocampus and molecular layer of cerebellum).
 CC -!- DEVELOPMENTAL STAGE: Expression increases from low levels at birth
 CC to high levels at 3-4 weeks before dropping slightly in adulthood.
 CC Expressed in the cortex and the molecular layer of the cerebellum
 CC at postnatal day 7. Isoform 2 expression does not change during
 CC development of both cortex and cerebellum. Isoform 4 expression
 CC decreases significantly during development of cortex but not
 CC cerebellum.
 CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.
 CC -!- SIMILARITY: Contains 7 ANK repeats.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF102855; AAD04569.2; -
 CC EMBL; AF131951; AAD29417.1; ALT_INIT.
 CC EMBL; AF159046; AAD42975.1; -
 CC EMBL; AF141904; AAF02498.1; ALT_INIT.
 CC HSSP; P00519; IABL.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001660; SAM.
 CC InterPro; IPR001452; SH3.
 CC Pfam; PF00023; ank; 7.
 CC Pfam; PF00595; PDZ; 1.
 CC Pfam; PF00536; SAM; 1.
 CC Pfam; PF00018; SH3; 1.
 CC PRODOM; PD000066; SH3; 1.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00228; PDZ; 1.
 CC SMART; SM00454; SAM; 1.
 CC SMART; SM00326; SH3; 1.
 CC PROSITE; PSS0088; ANK_REPEAT; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
 CC PROSITE; PSS0106; PDZ; 1.
 CC PROSITE; PSS0002; SH3; 1.
 CC PROSITE; PSS0105; SAM_DOMAIN; 1.
 CC KW ANK repeat; SH3 domain; Repeat; Alternative splicing.
 CC FT REPEAT 195 210 ANK 1.
 CC FT REPEAT 212 245 ANK 2.
 CC FT REPEAT 246 278 ANK 3.
 CC FT REPEAT 279 312 ANK 4.
 CC FT REPEAT 313 345 ANK 5.
 CC FT REPEAT 346 378 ANK 6.
 CC FT REPEAT 379 395 ANK 7.
 CC FT DOMAIN 554 613 SH3.
 CC FT DOMAIN 563 757 PDZ.
 CC FT DOMAIN 2104 2167 SAM.
 CC FT DOMAIN 929 932 POLY-PRO.
 CC FT DOMAIN 1010 1015 POLY-HIS.
 CC FT DOMAIN 1022 1027 POLY-HIS.
 CC FT DOMAIN 1194 1199 POLY-GLY.
 CC FT DOMAIN 1850 1860 POLY-PRO.
 CC FT VARSPPLIC 1 613 Missing (in isoform 2).
 CC FT VARSPPLIC 615 654 /FTid=VSP_006072.
 CC FT VARSPPLIC 615 654 SQGRQESRDKARLRFHRTVGSYDSFDPSPSLIDGIDSG
 CC FT -> MALSAVGCGPGGALPQPPFALSSSWPALGFRRSVMY
 CC FT IY (in isoform 2).
 CC FT /FTid=VSP_006073.
 CC FT Missing (in isoform 3).
 CC FT /FTid=VSP_006074.
 CC FT Missing (in isoform 4).
 CC FT /FTid=VSP_006075.

```

FT VARSPLIC 1930 1943 LSEDSQTSLLSKPS -> QYRIWVKSSDFGDF (in isoform 5)
FT /FTID=VSP 006076.
FT Missing (in isoform 5).
FT VARSPLIC 1944 2167 /FTID=VSP 006077.
FT S -> T (IN REF. 1).
FT CONFLICT 1141 1141 S -> N (IN REF. 2).
FT CONFLICT 1174 1174 S -> N (IN REF. 2).
FT CONFLICT 1246 1246 R -> K (IN REF. 1).
FT CONFLICT 1323 1323 A -> T (IN REF. 1).
FT CONFLICT 1331 1331 S -> D (IN REF. 1).
FT CONFLICT 1726 1726 S -> N (IN REF. 2).
SQ SEQUENCE 2167 AA; 226333 MW; 3F478B5A7B18BA86 CRC64;

Alignment Scores:
Pred. No.: 6 85 Length: 2167
Score: 83.00 Matches: 34
Percent Similarity: 36.79% Conservative: 5
Best Local Similarity: 32.08% Mismatches: 34
Query Match: 9.47% Indels: 33
DB: 1 Gaps: 5

US-10-057-510-1 (1-486) x SHK1_RAT (1-2167)
QY 195 CATGGAGCTGTGAGAGGCGCTGAGCTGAGAGTGTGACGAGCCCTGGAGGAGGGCA 254
Db HisAlaArgHisGluProValLeuArgLeu-----TrpGlyAspProAlaArgGlu 1425
QY 255 GAGGCTAC---CGAGGCTGAGGATGAGATGACGAGCCCTGGATNGT-----CCCCAG 305
Db LeuGlyTyzArgAlaGlyLeuGlySerGinGluLysAlaLeuThrAlaSerProAla 1445
QY 306 ACTCTCAGGACATTCGCCAGTCAAGGGTTGAGCCACNAGNCTTGCTCATATGCGATG 365
Db AlaArgArgSerLeu-----
QY 1446 AlaArgArgSerLeu-----1450
QY 366 AAGCGGAACTGTGATAGAGCCCTCCCTGTTGTGCTCGCCCTGCTGTTCTGGAA 425
Db -----LeuHisArgLeuProPro-ThrAlaProGlyValGlyProLeuLeuLeu 1467
QY 426 AACTGGTNTAGNCCCTC-----TGCTTTTNGCAA 458
Db nLeuGlyProGluProProThrProHisProGlyValSerLyAlaTyrArgThrAlaAl 1487
QY 459 CCTGGAAGAGGAA 474
Db aProGluGluProGlu 1492

RESULT 10
ID_KCH2_RABIT STANDARD; PRT; 1161 AA.
AC Q8WNY2; 002731; 019119; 097586; Q9TV06;
DT 28-FEB-2003 (Rel. 41, Created)
DT 15-MAR-2004 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Potassium voltage-gated channel subfamily H member 2 (Ether-a-go-go
DE related gene potassium channel 1) (ERG1) (RERG) (ra-erg) (Ether-a-go-
DE go related protein 1) (Eag related protein 1).
DE KCHN2 OR ERG.
GN Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Wicheel H.J., Hancox J.C., Levi A.J., Meech R.W.;
RT "RERG - rabbit ventricular ERG K+ channel subunit.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 411-571 FROM N.A. (ISOFORM 1/2).
RA MEDLINE=97164986; PubMed=9012748;
RP Wymore R.S., Gintant G.A., Wymore R.T., Dixon J.E., McKinnon D.,
RA Cohen I.S.;
RT "Tissue and species distribution of mRNA for the Ikr-like K+ channel,"

```

```

RT ERG."
RL CIRC. Res. 80:261-268(1997).
CC FUNCTION: Pore-forming (alpha) subunit of voltage-gated inwardly
CC rectifying potassium channel. Channel properties are modulated by
CC cAMP and subunit assembly. Mediates the rapidly activating
CC component of the delayed rectifying potassium current in heart
CC (IKr) (By similarity).
CC SUBUNIT: The potassium channel is probably composed of a homo- or
CC heterotetrameric complex of pore-forming alpha subunits that can
CC associate with modulating beta subunits. Heteromultimer with
CC KNH6/ERG2, KCNH7/ERG3, KCNE1 and KCNE2 (By similarity).
CC SUBCELLULAR LOCATION: Integral membrane protein.
CC ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8WNY2-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8WNY2-2; Sequence=VSP_000971;
CC TISSUE SPECIFICITY: Detected in heart, both in atrium and in left
CC ventricle.
CC DOMAIN: The segment S4 is probably the voltage-sensor and is
CC characterized by a series of positively charged amino acids at
CC every third position.
CC PM: Phosphorylated on serine and threonine residues (By
CC similarity).
CC SIMILARITY: Belongs to the potassium channel family. H (Eag)
CC subfamily.
CC SIMILARITY: Contains 1 cyclic nucleotide-binding domain.
CC SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
CC SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U87513; AAB68612.1; -.
CC EMBL; AF068736; AAC99425.1; -.
CC EMBL; AF105061; AAD39357.1; -.
CC EMBL; U75212; AAC48723.1; -.
CC InterPro; IPR000595; cNMP binding.
CC InterPro; IPR003967; Erg_channel.
CC InterPro; IPR005821; Ion_trans.
CC InterPro; IPR001622; K-channel_pore.
CC InterPro; IPR005820; M-channel_nlg.
CC InterPro; IPR001610; PAC.
CC InterPro; IPR000700; PAS-assoc_C.
CC InterPro; IPR000014; PAS_domain.
CC Pfam; PF00027; cNMP_binding; 1.
CC Pfam; PF00520; Ion_trans; 1.
CC Pfam; PF00785; PAC; 1.
CC PRINTS; PR01470; ERGCHANNEL.
CC SMART; SM00100; cNMP; 1.
CC SMART; SM00086; PAC; 1.
CC PROSITE; PS00888; cNMP_BINDING_1; FALSE_NEG.
CC PROSITE; PS00889; cNMP_BINDING_2; FALSE_NEG.
CC PROSITE; PS00042; cNMP_BINDING_3; 1.
CC PROSITE; PS00112; PAS; 1.
CC PROSITE; PS00113; PAC; 1.
CC Transport; Ion transport; 1.
CC Potassium channel; Potassium; Potassium transport; Transmembrane;
CC Phosphorylation; Glycoprotein; Multi-pass transmembrane; Alternative splicing.
CC DOMAIN 1 405 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 406 426 SEGMENT S1 (POTENTIAL).
CC TRANSMEM 453 473 SEGMENT S2 (POTENTIAL).
CC DOMAIN 474 497 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 498 518 SEGMENT S3 (POTENTIAL).
CC TRANSMEM 523 543 SEGMENT S4 (POTENTIAL).
CC DOMAIN 544 549 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 550 570 SEGMENT S5 (POTENTIAL).

```



```
Percent Similarity: 38.46% Conservative: 25
Best Local Similarity: 20.98% Mismatches: 55
Query Match: 9.82% Indels: 33
DB: 1 Gaps: 4

US-10-057-510-1 (1-486) x SPRA_CABEL (1-1311)
QY 423 CAGAACAGAGGGCCAGGACACACACAGGAGGGGGCTTCT----- 382
Db 921 ProglutylsArgLysGlnSerSerGluGluProAlaSerAspProGluLeuPheGly 940
QY 381 TATCAGAGTCCCTTCATGCGATATGAGCAGCAAGNCCCTGCTCGAACCCCTTGACTGG 322
Db 941 SerSerSerGlnProThrArgGlnLeuSerGluArgAlaThrArgAsnArgIleAsnTyr 960
QY 321 GCAATGCTCTGAGAGTCTGGGACNATCCAGGGGCTGATCTTCATCCTCAGCCTGCGT 262
Db 961 SerLeuLeuSerLysAsnGlySerGlyLysProThrProSerThrSerSerAlaAsnLeu 980
QY 261 AGCCTTCTGCTCCCTCCAGGCTGCTCCATTCCTTCAGCTCCAGGCTCCTCCACAC 202
Db 981 GluLysLeuAlaGlySerSerGlyLysSerGluSerProGluProAspGluSer 1000
QY 201 GTCCATGCTCTCCAC-----CTGCCACATGGANACAG----- 169
Db 1001 ValGluValSerHisTrpLysIleArgThrPheLeuArgSerGluTyrGlyValLysGlu 1020
QY 168 -----CGTGCCCCACAGCCAGGAGGACACAGAGCTGGCAAGAG 124
Db 1021 SerLeuLysCysProAspCysProTyrLysSerSerGluProAspValLeu----- 1037
QY 123 AACAGCAGTTAGACGAGGGGCAACATGGAGAGGAGTATCTATATGAGCTTGGGCAT 64
Db 1038 -----GluLysHisArgTyrTyrHisMetThrLys 1047

RESULT 12
DUFF_PAPHA
ID DUFF_PAPHA STANDARD; PRT; 336 AA.
AC Q95LIG5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Duffy antigen/chemokine receptor.
GN FY OR DARC
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
OX NCBI_TaxID=9557;
RN [1]
RP SEQUENCE FROM N.A.
RA Tournamille C., Blancher A., Le Van Kim C., Gane P., Apoil P.A.,
RA Nakamoto W., Cartron J.-P., Colin Y.;
RT "Sequences, evolution and ligand binding properties of mammalian duffy
RT antigen/receptor for chemokines";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Non-specific receptor for many chemokines (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family Duffy of G-protein coupled
CC receptors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/ebis-sib.ch/
CC or send an email to license@sib-sib.ch).
```

```
CC EMBL; AF303532; AAL09324.1; -
DR InterPro; IPR005384; Duffyantigen.
DR PRINTS; PR01559; DUFFYANTIGEN.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 63 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 64 84 1 (POTENTIAL).
FT DOMAIN 85 95 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 96 116 2 (POTENTIAL).
FT DOMAIN 117 126 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 127 147 3 (POTENTIAL).
FT DOMAIN 148 166 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 167 187 4 (POTENTIAL).
FT DOMAIN 188 207 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 208 228 5 (POTENTIAL).
FT DOMAIN 229 244 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 245 265 6 (POTENTIAL).
FT DOMAIN 266 287 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 288 308 7 (POTENTIAL).
FT DOMAIN 309 336 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 16 16 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 27 27 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 33 33 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 336 AA; 35603 MW; 5BA5DAD812DB685 CRC64;

Alignment Scores:
Pred. No.: 7.35 Length: 336
Score: 82.00 Matches: 55
Percent Similarity: 37.97% Conservative: 16
Best Local Similarity: 29.41% Mismatches: 56
Query Match: 9.36% Indels: 60
DB: 1 Gaps: 12

US-10-057-510-1 (1-486) x DUFF_PAPHA (1-336)
QY 42 GCGTGGAGAGTGGTGTGTAATAATGCCCAAGCTCATATAGATCTCTCTCTCCAT 98
Db 95 GlyTrpProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerI 111
QY 99 TGTGCCCCCTGCTTAACGTGCTTCTTCT-----TGCAGCT 137
Db 111 eValValProIleLeuAlaProGlyLeuGlyAsnThrArgSerSerAlaLeuCysSerLe 131
QY 138 TCGTGTTCCTCTCTG-----CTGTGGGG-----GC 164
Db 131 uGlyTyrCysValTrpTyrGlySerAlaPheAlaGlnAlaLeuLeuGlyCysHisAl 151
QY 165 ACGGCTGTNTCCA-----TGTGCAAGGTGGAGGCGATGGAGTGTGCGAGGAGCG 215
Db 151 aserLeuGlyProLysLeuGlyAlaAspGlnValProGlyLeu-ThrLeuGlyLeuServ 171
QY 216 CTGGAGCTGAAGGA-----ATGGACGAGCCCTGGGAGGAGGCGAGAGGCTAC 263
Db 171 aLgLyLeuTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGlyAla 190
QY 264 GCAGG-----CTGAGGATGAAGATGCAGCCCTGGATNGTCCCCAG 305
Db 191 SerGlyGlyLeuCysThrProValTyrSerMetGluLeuLysAlaLeuGlnAlaThrHis 210
QY 306 ACTCTCAGGACATTCGCCAGTCACAGGGTTCGAGGCACNAGNCCTT---GGCTCATATGC 362
Db 211 AlaValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLysGly 230
QY 363 ATGAAGGGGAACCTTCATAGAAGCCCTCCCTGG-----TTGTGTCTCTGGGCC 413
Db 231 LeuLysLysAlaLeuGlyMetGlyProGlyProIleMetAsnIleLeuTrpAlaTrp-P 250
QY 414 TCTGTCTCGAAAACACTGGTNTAGNCCCTCGTTTNTGCAAGACCTGAAGAAGGGA 473
Db 250 eilePheTrp-----Trp----- 254
QY 474 AATCCCCATGGG 486
```

```

Db      255 ----ProHisGly 257

RESULT.13
CP3E_CAVPO
ID CP3E_CAVPO STANDARD; PRT; 503 AA.
AC Q64417; Q64407;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 3A14 (EC 1.14.14.1) (CYPI1A14).
GN CYP3A14.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Hartley; TISSUE=Liver;
RX MEDLINE=98096351; PubMed=9434738;
RA Mori T., Itoh S., Ohgiya S., Ishizaki K., Kamataki T.;
RT "Regulation of CYP1A and CYP3A mRNAs by ascorbic acid in guinea
RT pigs.";
RL Arch. Biochem. Biophys. 348:268-277(1997).
CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC monooxygenases. In liver microsomes, this enzyme is involved in an
CC NADPH-dependent electron transport pathway. It oxidizes a variety
CC of structurally unrelated compounds, including steroids, fatty
CC acids, and xenobiotics.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- INDUCTION: P450 can be induced to high levels in liver and other
CC tissues by various foreign compounds, including drugs, pesticides,
CC and carcinogens.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D16363; BAA03865.1; -.
DR EMBL; D49731; BAA08568.1; -.
DR HSP; P14779; 1JP2.
DR InterPro; IPR001128; Cytochrome P450.
DR InterPro; IPR008072; EP450_CYP3A.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
FT METAL 442 442 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 121 121 S -> P (IN REF. 1).
FT CONFLICT 452 452 L -> I (IN REF. 1).
SQ SEQUENCE 503 AA; 58118 MW; 2348B0B3013FD2DD CRC64;

Alignment Scores:
Pred. No.: 7.58 Length: 503
Score: 82.00 Matches: 37
Percent Similarity: 36.99% Conservative: 17
Best Local Similarity: 25.34% Mismatches: 60
Query Match: 9.76% Indels: 32

DB:
US-10-057-510-1 (1-486) x CP3E_CAVPO (1-503)
QY 467 TCCTCAGGCTTTCGNAACCAACAGGGGCTGANGCAGTTTTCAGAACAGAGGCGCA 408
Db 29 SerHisGlyPhePheLysLysLeuGlyIle-----Pro 39
QY 407 GGACCAACACAGGAGGGGCTTCTTATGCAAGTTCCCTTCATGCC-----ATA 357
Db 40 GlyProLysPro-----LeuProLeuPheGlyAsnValLeuSer 52
QY 356 TGAGCCAAAGCCTNGTGGCTCGAACCTTGACTGGGCAATGCTCAGAGTCTGGGGACN 297
Db 53 TyrArgLysGlyMetTrpSerPheAspIle-----GluCysArgLysLysTyrGlyAsn 70
QY 296 ATCCAGGGGTGCATCTTCATCTCAGCCCTGCTAGCTTCTGCTCCTCCTCCAGGGCT 237
Db 71 MetTrpGlyLeuTyrAspGlyProGlnProValLeuAlaIleThrGluProAspMetIle 90
QY 236 CGTCCATTCTTCAGCTCCAG---CGCTCTCCACACGCTCCATGCTTCCACCTTGCCCA 180
Db 91 LysAlaValLeuValLysGluCysTyrSerValPheThrAsnArgSerLeuValPro 110
QY 179 CATGGANACAGCGCTGCTCCCTCCACAGCAGGAGGAAACAGAGCTGGCAAGAA---AGA 123
Db 111 ValGlyPheMetLysLysAlaValSerLeuSerGluAspGluLufTrpLysArgIleArg 130
QY 122 ACAGCAGTTAGACAGGGGCAACAATGGAGAGGAGTATCTATTAGAGCTTGGGCATT 63
Db 131 ThrGlnLeuSerProAsnPheThrSer-----GlyLys 141
QY 62 TTACACAGCAGCTTCCCA 45
Db 142 LeuLysGluMetPhePro 147

RESULT 14
CP3F_CAVPO
ID CP3F_CAVPO STANDARD; PRT; 503 AA.
AC Q64406;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 3A15 (EC 1.14.14.1) (CYPI1A15).
GN CYP3A15.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Hartley; TISSUE=Liver;
RX MEDLINE=98096351; PubMed=9434738;
RA Mori T., Itoh S., Ohgiya S., Ishizaki K., Kamataki T.;
RT "Regulation of CYP1A and CYP3A mRNAs by ascorbic acid in guinea
RT pigs.";
RL Arch. Biochem. Biophys. 348:268-277(1997).
CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC monooxygenases. In liver microsomes, this enzyme is involved in an
CC NADPH-dependent electron transport pathway. It oxidizes a variety
CC of structurally unrelated compounds, including steroids, fatty
CC acids, and xenobiotics.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- INDUCTION: P450 can be induced to high levels in liver and other
CC tissues by various foreign compounds, including drugs, pesticides,
CC and carcinogens.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D16363; BAA03865.1; -.
DR EMBL; D49731; BAA08568.1; -.
DR HSP; P14779; 1JP2.
DR InterPro; IPR001128; Cytochrome P450.
DR InterPro; IPR008072; EP450_CYP3A.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
FT METAL 442 442 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 121 121 S -> P (IN REF. 1).
FT CONFLICT 452 452 L -> I (IN REF. 1).
SQ SEQUENCE 503 AA; 58118 MW; 2348B0B3013FD2DD CRC64;

Alignment Scores:
Pred. No.: 7.58 Length: 503
Score: 82.00 Matches: 37
Percent Similarity: 36.99% Conservative: 17
Best Local Similarity: 25.34% Mismatches: 60
Query Match: 9.76% Indels: 32

```


Search completed: April 6, 2004, 17:56:04
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 6, 2004, 17:52:30 Search time 17 Seconds

(without alignments)
5499.892 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876

Sequence: 1 gcaacagaggtatnaaaaaat.....aaagggaantccccatggg 486

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.spool -DEV=xlp
-Q=/cgn2.1/USPTO.spool_p/US10057510/runat_06042004_141609_14737/app.query.fasta_1.647
-DB=PIR 78 -QFWT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10057510@cgn 1.1.38 @runat_06042004_141609_14737 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	84.5	10.1	1173	2 T31421	C-terminal domain-
C 2	84.5	10.1	1355	2 S40022	spalt protein - fr
C 3	84.5	10.1	2957	2 T33152	hypothetical prote
C 4	84	9.6	338	2 I52608	glycoprotein D - h
C 5	83.5	9.9	383	2 T46707	procephophoglyca
C 6	83.5	9.9	817	2 S51342	verprolin - yeast
C 7	83	9.9	1706	2 T51342	probable RNA helic
C 8	82.5	9.8	325	2 T96815	hypothetical prote
C 9	82.5	9.8	749	2 T08101	outer dynein arm d
C 10	82.5	9.8	1309	2 T19170	hypothetical prote
C 11	81	9.6	309	1 S03257	homeotic protein H
C 12	81	9.2	410	2 AG2968	hypothetical prote
C 13	81	9.2	410	2 D98314	hypothetical prote
C 14	81	9.6	772	2 T13078	KIAA0992 protein -

C 15	81	9.6	1791	2 T02345	hypothetical prote
C 16	80.5	9.6	217	2 B41256	sdr protein - chic
C 17	80.5	9.6	392	2 B48423	homeotic protein e
C 18	80	9.5	328	2 B88979	protein F37B4.10 l
C 19	80	9.5	360	1 A25732	inhibin alpha chai
C 20	80	9.5	540	1 F0MVHL	gas polypeptide -
C 21	79.5	9.5	268	2 A49303	homeotic protein C
C 22	79.5	9.5	401	2 A48423	engrailed homeodom
C 23	79.5	9.5	450	2 S00950	hypothetical prote
C 24	79.5	9.5	1239	2 G71266	probable ATP-depen
C 25	79.5	9.5	1309	2 T00078	probable RNA-direc
C 26	79	9.0	340	2 C69328	iron-sulfur cluste
C 27	78.5	9.0	107	2 T35634	hypothetical prote
C 28	78.5	9.0	178	2 G84357	probable acetyltra
C 29	78.5	9.3	1026	1 A40315	maternal effect pr
C 30	78.5	9.3	1734	2 A54602	microtubule-associ
C 31	78	8.9	154	2 T03956	hypothetical prote
C 32	78	9.3	834	2 S19155	serotonin receptor
C 33	77.5	9.2	173	2 A72450	hypothetical prote
C 34	77.5	8.8	865	2 T40170	hypothetical prote
C 35	77.5	9.2	1446	1 A45344	immediate-early pr
C 36	77	9.2	1060	2 S33641	homeotic protein z
C 37	76.5	8.7	275	2 T50578	hypothetical prote
C 38	76.5	9.1	301	2 T03914	hypothetical prote
C 39	76.5	8.7	652	2 D65215	acetate-CoA ligase
C 40	76.5	8.7	652	2 C91260	acetyl-CoA synthet
C 41	76.5	8.7	652	2 G86100	acetyl-CoA synthet
C 42	76.5	9.1	1017	2 T48452	hypothetical prote
C 43	76	9.0	654	2 A83553	membrane protein x
C 44	76	9.0	1032	2 T34433	hypothetical prote
C 45	76	9.0	2232	2 T34434	hypothetical prote

ALIGNMENTS

RESULT 1

T31421 C-terminal domain-binding protein rAI - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 29-Oct-1999 #sequence_revision 23-Oct-1999 #text_change 07-Dec-1999

C;Accession: T31421

R;Yuryev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebata, M.; Cord

Proc. Natl. Acad. Sci. U.S.A. 93: 6975-6980, 1996

A;Article: The C-terminal domain of the largest subunit of RNA polymerase II interacts wit

A;Reference number: Z21024; MUID:96293459; PMID:8692929

A;Accession: T31421

A;Status: preliminary; translated from GE/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 1-1173 <YUR>

A;Cross-references: EMBL:U49056; NID:G1438531; PID:G1438532; PIDN:AAC52657.1

A;Experimental source: hippocampus

Alignment Scores:

Pred. No.:	3.5	Length:	1173
Score:	84.50	Matches:	54
Percent Similarity:	39.43%	Conservative:	15
Best Local Similarity:	30.86%	Mismatches:	47
Query Match:	10.06%	Indels:	59
DB:	2	Gaps:	12

US-10-057-510-1 (1-486) x T31421 (1-1173)

QY	423	CCAGACAGAGGGCCAGGACACA-----ACCAGGAGGGGGCTTCTTATGCAAGTTC	370
DB	65	ProArgLeuArgAlaTrpThrGlyThrValSerProGlnSerHisAlaSerArg	84
QY	369	CCTTATGTCATATAGCCAGGNCCTTGCTGACACCTTGACTGGGCAATGCTCTGA	310
DB	85	ProAlaCys-----SerArgHisLeuLeu-----ThrLeuGly-----	95
QY	309	GAGTCTGGGACNAT-----CCAGGGGCTGCATCTTCATCCTCAGCCCTCGGT	262
DB	96	---ThrGlyAspGlyGlyProAlaProProAlaProSerSerGlySerSerPro	114

```
QY 261 AGCCTTCTGCTCCCTCCAGGCTCGTCCATTCCTTCAGCTCCAGCGCTCTCTCCACAC 202
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 SerProSerProSerSerSerProProProProProProProProProProProP 134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 201 GTCCATGCTTCCA-----CCTGCCACATGG 175
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 134 roProAlaLeuProAlaProArgPheAspIleTyrAspProPheHisProThrAspGluA 154
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 ANACAGCGGTCGCCCA-----CAAGCCAGAGGGGAACA 142
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 154 latSerProProProAlaProGluGlnLysTyrAspProPheGluAlaThrGlySerA 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 141 AC-----GAAGCTGCCAAGAAGAAGACAGC 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 snProSerSerGlyThrProSerProGluGluGluGluGluGluGluGluG 194
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 117 AGTTAGACAGGGG-----CAACAATGGGAAG-----GAGTATCTATTATGAGCTT 70
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 194 luGluGluGluGlyLeuSerGlnSerIleArgArgIleSerGluThrLeuAla----- 211
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 GGCATTTTAC-----ACGAGCATCTCCAGCTTTTCCA 35
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 212 --GlyIleTyrAspAspAsnSerLeuSerGlnAspPhePro 224
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 2
S40022
spalt protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C:Accession: S40022
R:Kuehlein, R.P.; Frommer, G.; Friedrich, M.; Gonzalez-Gaitan, M.; Weber, A.; Wagner-Ber
EMBO J. 13, 168-179, 1994
A>Title: spalt encodes an evolutionarily conserved zinc finger protein of novel structure
A:Reference number: S40022; MUID:94139659; PMID:7905822
A:Accession: S40022
A:Molecule type: DNA
A:Residues: 1-1355 <RUE>
A:Cross-references: EMBL:G15541; NID:G414105; PIDN:CAA53229.1; PID:G2598394
A>Note: the authors translated the codon GAC for residue 51 as Ala
A:Note: mRNA sequencing has also been done
C:Genetics:
A:Gene: sal; spalt
A:Cross-references: FlyBase:FBgn0004579
A:Introns: 51/1; 1329/3 1355/2
C:Keywords: zinc finger
Alignment Scores:
Pred. No.: 3.51 Length: 1355
Score: 84.50 Matches: 46
Percent Similarity: 39.29% Conservative: 20
Best Local Similarity: 27.38% Mismatches: 66
Query Match: 10.06% Indels: 36
DB: 2 Gaps: 7
US-10-057-510-1 (1-486) x S40022 (1-1355)
QY 441 GGNCTANGANCAGTTTCCAGAACAGAGGCCAGGACACACACACAGGAGGGGGTCTTCT 382
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1029 GlyValValAsnThrAsnProValArgProArgSerSerAlaSerSerHisGlyHisSer 1048
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 381 TAGTCAAGT-----TCCCTTCATGCTATGATGCGCAAGNCTNGTGCT 337
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1049 ValGlySerThrSerAlaProThrSerProSerValHisAlaSerSerGlnValIleLys 1068
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 336 CGAACCCCTTCACTGGGCAATGTCTCTGAGAGTCTGGGGAC-----NATCCAGGG 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1069 ArgSerSerSerProAlaArgSerGluAlaSerGlnGlyAlaLeuAspLeuThrProArg 1088
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 288 GTTGCA---TCTTCATCTCAGCCCTCGTAGCGTCTTGCCCTCCGCCAGGGCTCGTCC 232
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1089 AlaAlaProThrSerSerSerSerArgSerProLeuProLysGluLysProValSer 1108
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 231 ATTCCTTCACTCCAGCGC-----CTCTCCA 205
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1109 ProProSerLeuPro-ArgSerProSerGlySerSerHisAlaSerAlaAsnIleLeuTh 1128
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 204 CACGTCCATGCTTCCACC-----TTGCCACATGGANACAGCGGTGCCCC 160
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1128 rSerProLeuProProThrValGlyIleAspCysLeuProLysGlyLeuGlnHis---Hi 1147
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 159 CACAAGCCAGAGGAAACAAGAGCTGGCAAGAAAGAACAGCAGTGTAGACAGGGGCAAC 100
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1147 sLeuGlnGlnGlnHisGlnHisLeuMetGlnGlnGlnAlaValaAlaAlaA 1167
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 99 AATCGAAGAGGAGTATCTATTATGAGCTTGGGCAATTTATACCAGCACATTCACGCTT 40
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1167 aAlaGlnHisHis-----HisGlnGlnMetAlaAlaLe 1179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 39 TTCATAAACAATGGAACATTT 18
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1179 uAspGlnHisGlnGlnLeu 1186
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 3
T33152
Hypothetical protein T04D1.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33152
R:Davidson, S.; Wohldmann, P.
Submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid T04D1.
A:Reference number: Z21292
A:Accession: T33152
A>Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2957 <DAV>
A:Cross-references: EMBL:AF067617; PIDN:AAC17559.1; GSPDB:GN00019; CESP:T04D1.4
A:Experimental source: strain Bristol N2; Clone T04D1
C:Genetics:
A:Gene: CESP:T04D1.4
A:Map position: 1
A:Introns: 122/3; 293/3; 515/3; 1205/2; 1577/1; 2221/3; 2776/1; 2864/3
Alignment Scores:
Pred. No.: 3.53 Length: 2957
Score: 84.50 Matches: 39
Percent Similarity: 40.44% Conservative: 16
Best Local Similarity: 28.68% Mismatches: 52
Query Match: 10.06% Indels: 30
DB: 2 Gaps: 5
US-10-057-510-1 (1-486) x T33152 (1-2957)
QY 420 GAACAGAGGGCCAGGACACACACAGGAGGGGGCTTCTTATGGAAGTCCCTTCATGC 361
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2377 GlnGlnAlaGlnProSerThrSerLysGlyAlaSer---SerSerGlnSerSerAla 2395
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 360 CATATGAGCCAGNCTNGTGGCTCGAACCCCTTGACTGGGCAATG----- 316
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2396 GlnAlaAsnGlnAlaAlaGlnAlaGlnAlaAlaValaAlaGlnValaAla 2415
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 315 -----TCTCAGAGTCTTGGGACNATCCAGGGGCTGCATCTTCATCTCCAGCC 268
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2416 AlaAlaSerAlaSerThrSerSerAlaAlaGlnAlaValaGlnGlnGlnVal 2435
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 267 CTGGTAGCTTCTG----- 253
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2436 LeuGluAlaLeuLeuAlaMetSerMetAsnProAlaAlaLeuAlaLeuThrGlyThr 2455
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 252 CCCTCTCTCCAGGGGCTGCTCCATTCCTTCAGCTCCA-----GCGCTTCTCCACAC 202
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2456 SerSerSerThrGlyThrSerThrProSerAlaProLysLysSerLysProAlaProGln 2475
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 201 GTCCATGCTTCCACTTGCACATG---GANACAGCGGTGCCCCCAAGCCAGAGGG 146
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

192 AlSerGlyGlyLeuCysThrLeuLeuTySerThrGluLeuLysAlaLeuGlnAlaThr 211
303 CAGACTCTCAGGACATTGCCCGCAGTCAAGGGTTCGAGCCACNAGNCTT---GGCTCATAT 359
212 HisThrValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLys 231
360 GGCATGAAGGGGAACATTGCATTAAGAACCCCTCCCTGG-----TTGTGTCCTGG 410
232 GlyLeuLysLysAlaLeuGlyMetGlyProGlyProTrpMetAsnIleLeuTrpAlaTrp 251
411 CCCTCTGTTCTGGAAAACTGGTNCNTAGNCCCTCGTGTTTTTCGAAGCCCTGAAGAAAG 470
252 -PheIlePheTrp-----Trp----- 256
471 GGAAANTCCCCATGGG 486
257 -----ProHisGly 259
RESULT 5
T46707
proteophosphoglycan, membrane-associated [imported] - Leishmania major (fragment)
C:Species: Leishmania major
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 11-May-2000
C:Accession: T46707
R;Ilig, T.; Montgomery, J.; Stierhof, Y.D.; Handman, E.
J. Biol. Chem. 274, 31410-31420, 1999
A;Title: Molecular cloning and characterization of a novel repeat-containing Leishmania
atidylinositol anchor.
A;Reference number: 223133; MUID:20002668; PMID:10531342
A;Accession: T46707
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-383 <Ilig>
A;Cross-references: EMBL:AJ243460; PIDN:CAB46680.1
A;Experimental source: strain LRC-L137
C;Genetics:
A;Gene: pp91
C;Function:
A;Description: may be important for direct host-parasite interactions

[illegible]

RESULT 6

SS1342
verprolin - yeast (Saccharomyces cerevisiae)
N:Alternate names: prolin-rich protein VRP1; protein L8300.13; protein YLR337c
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1995 #sequence revision 11-Aug-1995 #text_change 23-Mar-2001
C:Accession: S51342; S39626; S57435
R:DNA, Z.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of *S. cerevisiae* cosmid 8300.
A:Reference number: S51339
A:Accession: S51342
A:Molecule type: DNA
A:Residues: 1-817 <DUZ>
A:Cross-references: EMBL:U9028; NID:G609380; PID:G609392; MIPS:YLR337c
R:Donnelly, S.F.H.; Focklington, M.J.; Pallotta, D.; Orr, E.
Mol. Microbiol. 10, 585-596, 1993
A:Title: A proline-rich protein, verprolin, involved in cytoskeletal organization and cell division in *S. cerevisiae*
A:Reference number: S39626; MUID:95058201; PMID:7968536
A:Accession: S39626
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-307, 'R', 309-349, 'R', 351-688, 'B', 690-709, 'HURWIPVPLIAPVKTLNNGYFLQVDRRCNWTSLI
A:Cross-references: EMBL:Z26645; NID:G414785; PID:CA81388.1; PID:G439289
R:Munn, A.L.; Stevenson, B.J.; Geli, M.I.; Riezman, H.
submitted to the EMBL Data Library, June 1995
A:Description: ends, end6, and end7: mutations that cause actin delocalization and block cell growth
A:Reference number: S57435
A:Accession: S57435
A:Molecule type: DNA
A:Residues: 1-162, 'F', 164-817 <MUN>
A:Cross-references: EMBL:X87806; NID:G871534; PID:G871535
A:Experimental source: strain W303
C:Genetics:
A:Gene: SGD:VRP1; MDP5; END5
A:Cross-references: SGD:S0004329; MIPS:YLR337c
A:Map position: 12R
F:5-14/Region: proline-rich
F:77-85/Region: proline-rich
F:114-180/Region: proline-rich
F:216-245/Region: proline-rich
F:305-336/Region: proline-rich
F:349-357/Region: proline-rich
F:372-382/Region: proline-rich
F:396-406/Region: proline-rich
F:421-445/Region: proline-rich
F:518-528/Region: proline-rich
F:567-577/Region: proline-rich
F:608-621/Region: proline-rich
F:649-661/Region: proline-rich
F:678-685/Region: proline-rich
F:704-710/Region: proline-rich

Alignment Scores:
Pred. No.: 4.39 Length: 817
Score: 83.50 Matches: 23
Percent Similarity: 48.57% Conservative: 11
Best Local Similarity: 32.86% Mismatches: 15
Query Match: 9.94% Indels: 21
DB: 2 Gaps: 2

US-10-057-510-1 (1-486) x S51342 (1-817)

QY 294 CCAGGGCTCATCTTCATCTCAGCCCTCGTAGCTTCG----- 253
Db 406 ProGlyAlaPheSerThrSerAlaLeuSerAlaSerSerTleProLeuAlaProLeu 425
QY 252 -----CCCTCTCCAGGGCTCGTCCATCTCTTCAGCTCCAGCGCTCTCCCA 205
Db 426 ProProProProProProSerValAlaThrSerValProSerAlaProProProPro 445
QY 204 CAGTCCTATGCTTCCACCTTGCCACATGGANACAGCGGTGCCCCCAAGCCAGAGGGA 145

Db 446 ThrLeu-----ThrThrAsnLysProSerAla 454

QY 144 ACAACGAGCTGGCAAGAAAGACAGCAGT 115

Db 455 SerSerLysGlnSerLysValSerSerSer 464

RESULT 7

B75633
probable RNA helicase - *Deinococcus radiodurans* (strain R1)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: B75633
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.
Science 286, 1571-1577, 1999
S: Smith, H.O.; Venter, J.C.; Fraser, C.M.
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: B75633
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1706 <WHI>
A:Cross-references: GB:AS001826; NID:G6460827; PID:AAF12564.1; PID:G6460860; TIGR:DRB0:
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRB0135
A:Map position: megaplasmid
A:Genome: plasmid
A:Note: plasmid MPI

Alignment Scores:
Pred. No.: 4.95 Length: 1706
Score: 83.00 Matches: 40
Percent Similarity: 33.52% Conservative: 20
Best Local Similarity: 22.35% Mismatches: 52
Query Match: 9.88% Indels: 67
DB: 2 Gaps: 6

US-10-057-510-1 (1-486) x B75633 (1-1706)

QY 429 AGTTTCCAGAACAGAGGGCCAGGACACACACAGGAGGGGGTCTTATGCAAGTTCC 370

Db 818 ThrLeuProArgAspGluGlyArgSerThrGlyArgLeuSerLeuAlaGlnGlyAla 837

QY 369 CCTTCATGCCATATGAGCCAAAGCCTGTGGCTCGAACCTTGACTGGGCAATCTCTGA 310

Db 838 ValGlyArgTyrLeuThrArg-----ProGluThrLeuAlaTrpSerGlnAlaLeu 854

QY 309 GAGTCTGGGAC----- 298

Db 855 LysArgAspValAlaAlaLeuHisAspLeuLeuAlaLeuSerGluPheLeu 874

QY 297 -----NAT 295

Db 875 GluGlnValGluProGlyAlaTyrGlnLeuLysGlyThrSerPheThrTrpHisAlaGly 894

QY 294 CCAGGGCTGATCTTCATCTCAGCCCTCGTAGCTTCGCTCCAGGGCTCG 235

Db 895 ProGlyThrGluProSerValAlaLeuArgValLeuArgPro----- 909

QY 234 TCCATCTCTCAGCTCCAGCGCTCTCCACACCTCCATCCCTCCACCTTCCACATGG 175

Db 910 -----AlaGlySerGluAlaProArgValAsnSerPheLeuSerLeuTyr 925

QY 174 ANACAGCGTCCCG-----CACAGCCAGAGGGAACACAGAACGCTG--- 133

Db 926 GlnGlnProAlaAlaGluPheArgAspLeuHisGlyAlaGluHisThrAlaGlnIleArg 945

QY 132 -----GCAAGAAAGAACAGCAGTTAGAGCAGGGGCA----- 102

Db 946 AlaGluGluArgGluLysArgGluGlu-AlaPheArgAlaGlyThrLeuProAlaLeuPhe 965

QY 101 -----ACAATGGAGAAAGGAGTATCTATTATGAGCTTGGCGCATTTTACAC 57
|||||
Db 965 eCysserProThrMetGluLeuGlyValAspLeuSerAspLeuValHis 993
|||||
RESULT 8
H96815
hypothetical protein F9K20.26 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96815
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.;
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96815
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-325 <STO>
A:Cross-references: GB:AB005173; NID:G3834322; PIDN:AAC83038.1; GSPDB:GN00141
C:Genetics:
A:Gene: F9K20.26
A:Map position: 1

Alignment Scores:
Pred. No.: 5.47 Length: 325
Score: 82.50 Matches: 41
Percent Similarity: 35.57% Conservative: 12
Best Local Similarity: 27.52% Mismatches: 37
Query Match: 9.82% Indels: 59
DB: 2 Gaps: 8

US-10-057-510-1 (1-486) x H96815 (1-325)

QY 394 GGAGGGGCTTTATGCAAGTTCCTTCATGCAATGAGCCAAAGNCCTGTCG 335
|||||
Db 84 GlyGlySerAlaThrAlaSerProCysSerTyrGlnProSerProCysAlaSer 103
|||||
QY 334 ---AACCCTTGACTGGCAATCTCTGAGAGCTCGGGACNATCCAGGGCTGCATCTC 278
|||||
Db 104 TyrAsnPro-----SerProGlySerSerAsn 112
|||||
QY 277 ATCTCAGCCCTCGTAGCCTTCT-----GCCCTC 248
|||||
Db 113 PheMetSerProAlaSerSerSerPheAlaAsnLeuThrSerGlyAspGlyGlnSerLeu 132
|||||
QY 247 CTCACAGGCTCGTCCAT-----TCCTTCAGCTCCAGCGCTCTCCACAGT 200
|||||
Db 133 IleProTrpLeuLeuHisLeuSerThrThrSerSerSerAlaSerSerSerArg 152
|||||
QY 199 -----CCATGCTTCCTCCACCTTGC 182
|||||
Db 153 LeuProAsnTyrLeuTyrIleProGlyGlySerIleSerAlaProValThrProLeu 172
|||||
QY 181 CACATGGANACGCGTGGCCCAACAGCCAGAGGACACAGAGCTGGCAAGAA 122
|||||
Db 173 SerSerProThrAlaArgThrPro-----ArgMetAsnThrAsp-----Trp 186
|||||
QY 121 CAGCAGTATAGACAGGGGCAACATGGAGAAAGGATATCTATTATGAGCTTGGCATTT 62
|||||
Db 187 GlnGlnLeuAsnAsn-----SerPhe 193
|||||
QY 61 TACACAGAGCTTCCAGGCTTTTCA 35
|||||
Db 194 PheValSerSerThrProProSerPro 202
|||||

RESULT 9

T08101
outer dynein arm docking complex protein ODA3 - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 08-Oct-1999
C:Accession: T08101
R:Koutoulis, A.; Pazour, G.J.; Wilkerson, C.G.; Inaba, K.; Sheng, H.; Takada, S.; Witmar
J. Cell Biol. 137, 1069-1080, 1997
A:Title: The Chlamydomonas reinhardtii ODA3 gene encodes a protein of the outer dynein
A:Reference number: Z16354; MUID:97311076; PMID:9166407
A:Accession: T08101
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-749 <KOU>
A:Cross-references: EMBL:AF001309; NID:G2150040; PIDN:AAC49732.1; PID:G2150041
C:Genetics:
A:Gene: ODA3
C:Function:
A:Description: may play an important role in the precise positioning of the outer dynein

Alignment Scores:
Pred. No.: 5.51 Length: 749
Score: 82.50 Matches: 35
Percent Similarity: 34.56% Conservative: 12
Best Local Similarity: 25.74% Mismatches: 36
Query Match: 9.82% Indels: 53
DB: 2 Gaps: 7

US-10-057-510-1 (1-486) x T08101 (1-749)

QY 411 GCCAGGACCACACAGGAGGGGCTTCTTATGCAAGTTCCTTCATGCCATATGAGC 352
|||||
Db 485 AlaSerAlaGlyAlaArgGlySerAlaHisThrProAspArgAsn----- 502
|||||
QY 351 CAAGNCCTGCTGCAACCTTGACTGGCAATGCTCTGAGAGTCTGGGACNATCCA 292
|||||
Db 503 -----LysArgGlyProAlaThr 508
|||||
QY 291 GGGCTGCATCTTCATCCTCAGCCCTCGTAGCTGCTGCTCCCTCCAGGCTCGTCC 232
|||||
Db 509 GlySerArgSerGlnSerProAla-----LeuValProHisSerProAlaGlyAsp 525
|||||
QY 231 ATTCCTTCAGCTCCA-----GGCCTCTCCACACGCTCCATGCCCTTCCACCTTGC 181
|||||
Db 526 LysProSerSerProLeuHisGlyThrSerProGluHisGlyHis----- 540
|||||
QY 180 ACATGGANACGCGTGGCCCAACAGCCAGAGGGACACAGAGCTGGCAAGAAAGAAC 121
|||||
Db 541 -----GluProIle-----ProGluGlyAlaGluLeuAla----- 551
|||||
QY 120 AGCAGTTAGAGCAGGGGCAACATGGAGAAAGGAGTATCTATTATGAGCTTGGG----- 67
|||||
Db 552 -----GlyGluAlaGluLeuValSerProLeuGlyAlaAsp 563
|||||
QY 66 -----CATTTACACAGCACTTCCACGCTT 40
|||||
Db 564 GlyAsnThrIleGluAspGluHisPheProGluLeuProGluLeu 579
|||||

RESULT 10

T19170
hypothetical protein C09H6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19170
R:Harris, B.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19083
A:Accession: T19170
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1309 <WIL>
A:Cross-references: EMBL:Z81466; PIDN:CAB03868.1; GSPDB:GN00019; CESP:C09H6.1
A:Experimental source: clone C09H6
C:Genetics:

A:Gene: CESP:C09H6.1
A:Map position: 1
A:Introns: 107/3; 434/3; 478/3; 564/2; 1020/1; 1240/3

Alignment Scores:
Pred. No.: 5.53 Length: 1309
Score: 82.50 Matches: 30
Percent Similarity: 38.46% Conservative: 25
Best Local Similarity: 20.98% Mismatches: 55
Query Match: 9.82% Indels: 33
DB: 2 Gaps: 4

US-10-057-510-1 (1-486) x T19170 (1-1309)

QY 423 CCAGAACAGAGGGCCAGACACACACAGGAGGGGGCTTCT----- 382
Db 921 ProGluLysLysArgLysGlnSerSerGluLupProAlaSerAspProGluLeuPheGly 940
QY 381 TATGCAGTTCCCTTCATGCCATATGAGCCAGAGNCTGTGGCTGCAACCTTGACTGG 322
Db 941 SerSerSerGlnProThrA:GlnLeuSerGluArgAlaThrArgAsnArgileAsnTyr 960
QY 321 GCATGTCCCTGAGAGTCTGGGACNATCCAGGGGCTGCATCTTCATCCAGCCCTGCGT 262
Db 961 SerLeuSerLysAsnGlySerGlyLysProThrProSerThrSerSerAlaAsnLeu 980
QY 261 AGCCTTTCGCTCTCCAGGCTGCTCCATTCCTTCAGCTCCAGGCTCTCCACAC 202
Db 981 GluLysLeuAlaGlySerSerGlyAlaSerSerGluSerProGluProAspGluSer 1000
QY 201 GTCCATGCTTCCAC-----CTGCCACATGGANACAG----- 169
Db 1001 ValGluValSerHisLysLysLysLysLysLysLysLysLysLysLysLysLys 1020
QY 168 -----CCGTCCCCCAACAGCCAGAGGAGAACAGAGCTGCGACAGAAAG 124
Db 1021 SerLeuLysCysProAspCysProTyrLysSerSerGluProAspValLeu----- 1037
QY 123 AACAGCATTAGACAGGGGACACATGGAGAGAGAGTATCTATTATGAGCTTGGGCAT 64
Db 1038 -----GluLysHisArgTyrThrHisMetThrLys 1047
QY 63 TTTACACCA 55
Db 1048 ThrThrPro 1050

RESULT 11

S09257
homeotic protein Hox A4 - chicken
N:Alternate names: homeotic protein Chox1-4
C:Species: Gallus gallus (chicken)
C>Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 21-Jul-2000
C:Accession: S09257; S10883
R:Saaki, H.; Yokoyama, E.; Kuroiwa, A.
Nucleic Acids Res. 18, 1739-1747, 1990
A:Title: Specific DNA binding of the two chicken deformed family homeodomain proteins,
A:Reference number: S09256; MUID:90245562; PMID:1970866
A:Accession: S09257
A:Molecule type: mRNA
A:Residues: 1-309 <SAS>
A:Cross-references: EMBL:X52670; NID:g63218; PIDN:CAA36896.1; PID:g63219
R:Scotting, P.J.; Hewitt, M.; Keynes, R.J.
Nucleic Acids Res. 18, 3999, 1990
A:Title: Isolation and analysis of chick homeobox cDNA clones.
A:Reference number: S10883; MUID:90326535; PMID:1973835
A:Accession: S10883
A:Molecule type: mRNA
A:Residues: 207-273 <SCO>
A:Cross-references: EMBL:X52747; NID:g63223; PIDN:CAB57949.1; PID:g6018426
C:Function:
A:Description: control of embryonic development by tissue- and stage-specific regulation
C:Superfamily: homeotic protein Hox D4; homeobox homology
C:Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation

F:210-266/Domain: homeobox homology <HOX>

Alignment Scores:
Pred. No.: 7.7 Length: 309
Score: 81.00 Matches: 31
Percent Similarity: 34.91% Conservative: 6
Best Local Similarity: 29.25% Mismatches: 31
Query Match: 9.64% Indels: 38
DB: 1 Gaps: 4

US-10-057-510-1 (1-486) x S09257 (1-309)

QY 341 TGGCTCGAACCTTGACTGGGCAATGCTCTGAG-----AGTCTGGG 300
Db 12 TyrileGluProLysPheProProCysGluGluTyrThrGlnHisSerGlySerAlaGly 31
QY 299 ACNATCCAGGGGCTGCATCTTCATCTCAGCCCTCGGTAGCTTCTGCTCTCCCTCCAGG 240
Db 32 SerSerAlaSerTyrHisProHisHis-----ProHisProHis 44
QY 239 GCTCGTCCATTCCTTCAGCTCCAGGCTCTCCACAGCTCCATGCTTCCACCTTGCCA 180
Db 45 AlaProPro--ProProProProProProHisLeuHisAlaHis----- 61
QY 179 CATGGANACAGCCGTGCCCCACAAAGCCAGAGGAGAACAAAGCTGCGAAGAAAGAAC 120
Db 62 -----ProGlyProAlaLeuProGluTyrPheProArgProArg 75
QY 119 GCAGTTAGAGCGGGGCAACATCGAGAGAGGATATCTATTATGAGCTTGGGCATTTTA 60
Db 76 -----GluProGlyTyrGlnA 81
QY 59 CACCAGCACTTCCC 46
Db 81 laProAlaAlaPro 85

RESULT 12

AG2968
hypothetical protein Atu3352 [imported] - Agrobacterium tumefaciens (strain C58, Dupont
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AG2968
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, J.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell.
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743133
A:Accession: AG2968
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <KUR>
A:Cross-references: GB:AB008689; PIDN:AAL44165.1; PID:g17741740; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3352
A:Map position: linear chromosome

Alignment Scores:
Pred. No.: 7.72 Length: 410
Score: 81.00 Matches: 37
Percent Similarity: 33.76% Conservative: 16
Best Local Similarity: 23.57% Mismatches: 47
Query Match: 9.25% Indels: 57
DB: 2 Gaps: 8

US-10-057-510-1 (1-486) x AG2968 (1-410)

QY 125 TTTCTTGCACCTTCGTTGTTCTCCCTGCTTGGGGGACGCTGTNTCCATGTCGCA 184
Db 10 PheLeuAlaSerValMetileProAlaGlyMetAlaGlyAlaThrAspLeuGluValThr 29

```

QY 185 AGGTGGAAGGCATGGACGTGTGGAGGAGCGCTGGAGCT----- 223
Db 30 HisTrp-----TrrhrSerGlyGluAlaAlaValAlaGluLeuAlaLysAla 47
QY 224 ---GAAGGAATGACGAGCCCTGGGAGGAGGCG----- 253
Db 48 PheAspAlaThrGlyAsnLysTrpValAspGlyAlaAlaLeuAlaGlySerGlyGlyThrAla 67
QY 254 -----AGAAAGGCTACGC----- 265
Db 68 ArgProIleMetIleSerArgIleThrGlyGlyAspProMetGlyAlaThrGlnPheAsn 87
QY 266 -----AGGCTGAGGATGAAGATGCGAGCCCTGGATNGTCCCGAGACTCTCAGGACATT 319
Db 88 HisGlyArgGlnAlaGluGluLeuValGlnAlaGlyLeuMetArgAspLeuSerAspVal 107
QY 320 GCCCAGTCAAGG-----TTTCGAGCCACNAGGN---CTTGGCTCATATGG 361
Db 108 Ala-AlaLysGlyAsnTrpLysGluValIleLysProAlaSerLeuLeuAspSerCysTh 127
QY 362 CATGAGGGGAACCTGCATAAGAAGCC-----CTCCCTCGTGTGGTCTCGGCC 412
Db 127 xLeAspGlyLysIleTy-CysAlaProValAsnIleHisSerTrpGlnTrpLeuTriple 147
QY 413 CTCT-----GTTCTGGAAAACCTGG 431
Db 147 uSerAsnAlaAlaPheLysLysAlaGlyValGluValProLysAsnTrp 163

RESULT 14
KIAA0992 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13078; T08742
R:Negase, T.; Ishikawa, K.; Suyama, M.; Kikuno, R.; Hirose, M.; Miyajima, N.; Tanaka,
DNA Res. 6, 63-70, 1999
A:Title: Prediction of the coding sequences of unidentified human genes. XIII. The comp
A:Reference number: Z17602; MUID:99245063; PMID:10231032
A:Accession: T13078
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-772 <NAG>
A:Cross-references: EMBL:AB023209; NID:94589627; PIDN:BAAY6836.1; PID:94589628
A:Experimental source: brain
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: T16471
A:Accession: T08742
A:Molecule type: mRNA
A:Residues: 628-772 <WAM>
A:Cross-references: EMBL:AL050093
A:Experimental source: adult uterus; clone DKFZp586L0518
C:Genetics:
A:Note: KIAA0992; DKFZp586L0518.1

Alignment Scores:
Pred. No.: 7.76 Length: 772
Score: 81.00 Matches: 35
Percent Similarity: 38.21% Conservative: 12
Best Local Similarity: 28.46% Mismatches: 40
Query Match: 9.64% Indels: 36
DB: 2 Gaps: 6

US-10-057-510-1 (1-486) x T13078 (1-772)
QY 486 CCCATGGGGANTTTCCCTTTTCAGGCTTTGCNAAAAACCCAGGGGNGNCTANGNCCAGT 427
Db 82 ProProAlaPheProGluLeuAlaAlaCysThrProProAlaSer----- 97
QY 426 TTTTCAGACAGAGGCGCCAGGACCAACAGGAGGGGGCTTCTTATGCAAGTCCCT 367
Db 98 ---ProGluProMetSerAlaLeuAlaSerArgSerAlaProAlaMetGlnSerSergly 116
QY 366 TCATGCATATGAGC-----CAAGNCTNGTGGCTCGAACCCCTTGACTGGGCAATG 316
Db 117 SerPheAsnTyrrAlaArgProLysGlnPheIleAlaAlaGlnAsnLeuGlyProAla--- 135
QY 315 TCCTGAGAGTCTGG---GACNATCCAGGGGCTGCATCTTCATCTCAGCCCTGCCTAGC 259
Db 136 -----SerglyHisGlyThrProAlaSerSerProSerSerSerSerProSer 152
QY 258 CTCTCG-----CCCTCCTCCAG--- 241

```

```

QY 185 AGGTGGAAGGCATGGACGTGTGGAGGAGCGCTGGAGCT----- 223
Db 30 HisTrp-----TrrhrSerGlyGluAlaAlaValAlaGluLeuAlaLysAla 47
QY 224 ---GAAGGAATGACGAGCCCTGGGAGGAGGCG----- 253
Db 48 PheAspAlaThrGlyAsnLysTrpValAspGlyAlaAlaLeuAlaGlySerGlyGlyThrAla 67
QY 254 -----AGAAAGGCTACGC----- 265
Db 68 ArgProIleMetIleSerArgIleThrGlyGlyAspProMetGlyAlaThrGlnPheAsn 87
QY 266 -----AGGCTGAGGATGAAGATGCGAGCCCTGGATNGTCCCGAGACTCTCAGGACATT 319
Db 88 HisGlyArgGlnAlaGluGluLeuValGlnAlaGlyLeuMetArgAspLeuSerAspVal 107
QY 320 GCCCAGTCAAGG-----TTTCGAGCCACNAGGN---CTTGGCTCATATGG 361
Db 108 Ala-AlaLysGlyAsnTrpLysGluValIleLysProAlaSerLeuLeuAspSerCysTh 127
QY 362 CATGAGGGGAACCTGCATAAGAAGCC-----CTCCCTCGTGTGGTCTCGGCC 412
Db 127 xLeAspGlyLysIleTy-CysAlaProValAsnIleHisSerTrpGlnTrpLeuTriple 147
QY 413 CTCT-----GTTCTGGAAAACCTGG 431
Db 147 uSerAsnAlaAlaPheLysLysAlaGlyValGluValProLysAsnTrp 163

RESULT 13
D98314
hypothetical protein AGR_L_2941 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: D98314
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: D98314
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK90038.1; PID:gi5160015; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2941
A:Map position: linear chromosome

Alignment Scores:
Pred. No.: 7.72 Length: 410
Score: 81.00 Matches: 37
Percent Similarity: 33.76% Conservative: 16
Best Local Similarity: 23.57% Mismatches: 47
Query Match: 9.25% Indels: 57
DB: 2 Gaps: 8

US-10-057-510-1 (1-486) x D98314 (1-410)
QY 125 TTTCTTCCAGCTTGTGTGTTCCCTCTGCTGTGGGGGACGCGCTGNTCCATGTGCA 184
Db 10 PheLeuAlaSerValMetIleProAlaGlyMetAlaGlyAlaThrAspLeuGluValThr 29
QY 185 AGGTGGAAGGCATGGACGTGTGGAGGAGCGCTGGAGCT----- 223
Db 30 HisTrp-----TrrhrSerGlyGluAlaAlaValAlaGluLeuAlaLysAla 47
QY 224 ---GAAGGAATGACGAGCCCTGGAGGAGGCG----- 253
Db 48 PheAspAlaThrGlyAsnLysTrpValAspGlyAlaAlaLeuAlaGlySerGlyGlyThrAla 67
QY 254 -----AGAAAGGCTACGC----- 265

```


Db 153 ProMetSerProThrProArgGlnPheGlyArgAlaProValProPheAlaGlnPro 172
Qy 240 -----GGCTGCTCCATTCTTCAGCTCCAGCGCTCCT 208
Db 173 PheGlyAlaGluProGluAlaProTyrGlySerSerProSerProProProPro 192
Qy 207 CCACAGGTC 199
Db 193 ProProVal 195

RESULT 15

T02345
hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R;Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
re, J.; White, S.; Teng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998.
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: Z14664
A:Accession: T02345
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C:Genetics:
A:Map position: 16
A:Introns: 1610/2; 1706/2
A:Note: KIAA0324

Alignment Scores:
Pred. No.: 7.81 Length: 1791
Score: 81.00 Matches: 52
Percent Similarity: 42.07% Conservative: 17
Best Local Similarity: 31.71% Mismatches: 50
Query Match: 9.64% Indels: 45
DB: 2 Gaps: 11

US-10-057-510-1 (1-486) x T02345 (1-1791)

Qy 464 TCAGGCTTTCNAAACACAGGGGNCNTANGANCAGTTTCCAGAACAGGGCCAGGA 405
Db 426 AlaGlyMetSerSerAsnGlnSerIleSerSerProValLeu--AspAlaValProArgT 445
Qy 404 CCACAAACAGGGAGGGGCTTTATGCAAGTTCCTCCATGATGAGCCAGNCC 345
Db 445 hrProSerArgGluArgSerSerSerAlaSerPro-----GluMetLysAspGlyL 463
Qy 344 TNGTGGCTCGAACCTTGACTGGGCAATGCTCTGAGAGTCTGGGACNATCCAGGGCT- 286
Db 463 eu---ProArgThrPro-----SerArgSerArgSerGlySerProGlyLeuA 480
Qy 285 -----GCATCTTCATCTCAGCCCTCGTAGCTTCTGCCCTCTCCAGGGC---TCGT 234
Db 480 rgaspGlySerGlyThrProSerArgHisSerLeuSerGlySerSerProGlyMetLysA 500
Qy 233 CCATTCCTTCAGTCCA-----GGCTCTCTCCACAGC 201
Db 500 spileProArgThrProSerArgGlyArgSerGluCysaspSerSerProGluPro-Lys 519
Qy 200 TCCATGCTTCCACCTTGCCACATGGANACAGCGCTGCCCCACACAGC----- 153
Db 520 AlaLeuProGlnThr---ProArgProArgSerProSerSerProGluLeuAsn 538
Qy 152 -----CAGAGGGAACACGAAGCTGGCAAGAA----- 126
Db 539 AsnLysCysLeuThrProGlnArgGlu---ArgSerGlySerGluSerSerValAspGln 557
Qy 125 -----AGACAGCAGGTT-----AGAGCAGGGGCAACAATGGAGAAA 90
Db 558 LysThrValAlaArgThrProLeuGlyGlnArgSerArgSerGlySerSerGlnGluLeu 577

Qy 89 GGAGTA 84
Db 578 AspVal 579

Search completed: April 6, 2004, 17:58:32
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 6, 2004, 17:55:25 ; Search time 42.5 Seconds
(without alignments)
6006.579 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876

Sequence: 1 gcacagcgatnaaaaaat.....aaagggaaantccccatggg 486

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Xgapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1071772 seqs, 262633353 residues

Total number of hits satisfying chosen parameters: 2143544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p_model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_P/US10057510/runat_06042004_141610_14799/app_query.fasta.1.647
-DB=published_applications_AA -QFM=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -L-STS=45 -DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10057510 @cgn 1 1 13 @runat 06042004_141610_14799
-NCPUS=6 -ICPU=3 -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	394	45.0	74	13	US-10-057-510-2	Sequence 2, Appli
2	387	44.2	247	15	US-10-131-410-116	Sequence 116, App
3	183.5	20.9	196	15	US-10-264-237-2428	Sequence 2428, Ap
4	94.5	11.2	19695	15	US-10-084-846A-3	Sequence 3, Appli
5	87.5	10.0	1725	14	US-10-029-388-27616	Sequence 27616, A
6	87.5	10.0	19723	15	US-10-084-846A-5	Sequence 5, Appli
7	86	10.2	163	9	US-09-738-973-41	Sequence 41, Appl
8	86	10.2	163	9	US-09-854-133-41	Sequence 41, Appl
9	86	10.2	163	14	US-10-144-649A-41	Sequence 41, Appl
10	86	10.2	378	14	US-10-156-761-9745	Sequence 9745, Ap
11	84.5	10.1	196	12	US-10-425-114-52720	Sequence 52720, A
12	84.5	10.1	250	12	US-10-425-114-53402	Sequence 53402, A
13	84.5	10.1	250	12	US-10-425-114-72216	Sequence 72216, A
14	84.5	10.1	261	12	US-10-425-114-63205	Sequence 63205, A
15	84	9.6	338	14	US-10-171-311-71	Sequence 71, Appl
16	84	9.6	338	14	US-10-225-567A-110	Sequence 110, Appl
17	83.5	9.9	817	15	US-10-369-493-1813	Sequence 1813, Ap
18	83	9.9	192	15	US-10-108-260A-4083	Sequence 4083, Ap
19	83	9.9	213	14	US-10-147-170A-5	Sequence 5, Appli
20	83	9.9	217	15	US-10-108-260A-4063	Sequence 4063, Ap
21	83	9.9	221	14	US-10-147-170A-6	Sequence 6, Appli
22	83	9.9	239	9	US-09-925-299-834	Sequence 834, App
23	83	9.9	239	10	US-09-925-299-834	Sequence 834, App
24	83	9.9	239	14	US-10-106-698-5808	Sequence 5808, Ap
25	83	9.9	254	15	US-10-334-143-29	Sequence 29, Appl
26	83	9.5	542	12	US-10-276-774-1386	Sequence 1386, Ap
27	82.5	9.8	165	12	US-10-425-114-45724	Sequence 45724, A
28	82.5	9.8	213	12	US-10-220-120-415	Sequence 415, App
29	82.5	9.8	597	15	US-10-369-493-3265	Sequence 3265, Ap
30	82	9.4	119	12	US-10-424-599-180662	Sequence 180662,
31	81	9.6	371	12	US-10-425-114-71775	Sequence 71775, A
32	80.5	9.2	116	12	US-10-424-599-222962	Sequence 222962,
33	80.5	9.6	228	12	US-10-425-114-68832	Sequence 68832, A
34	80	9.5	717	14	US-10-156-761-14833	Sequence 14833, A
35	80	9.5	1265	9	US-09-862-027-80	Sequence 80, Appl
36	80	9.5	1798	16	US-10-311-034-23	Sequence 23, Appl
37	79.5	9.5	192	12	US-10-425-114-68764	Sequence 68764, A
38	79.5	9.5	193	12	US-10-425-114-67400	Sequence 67400, A
39	79.5	9.5	651	15	US-10-346-144-2	Sequence 2, Appli
40	79.5	9.5	724	15	US-10-346-144-5	Sequence 5, Appli
41	79.5	9.5	1239	12	US-10-282-182A-76629	Sequence 76629, A
42	79	9.4	157	11	US-09-864-408A-1498	Sequence 1498, Ap
43	79	9.4	5546	12	US-10-296-734-1210	Sequence 1210, Ap
44	79	9.4	19608	15	US-10-084-846A-8	Sequence 8, Appli
45	78.5	9.0	243	14	US-10-050-704-283	Sequence 283, App

ALIGNMENTS

RESULT 1
US-10-057-510-2
; Sequence 2, Application US/10057510
; Publication No. US20020098580A1
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Yang, Meijia
; APPLICANT: Schulz, Vincent
; APPLICANT: Curagen Corporation
; TITLE OF INVENTION: MDM INTERACTING PROTEIN AND METHODS OF USE THEREOF
; FILE REFERENCE: 15966-524 MDM US
; CURRENT APPLICATION NUMBER: US/10/057,510
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: USSN 09/510,252
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: USSN 60/121,192
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: USSN 60/122,643
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 74
; TYPE: PRT

```
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Any X can be any amino acid.
US-10-057-510-2
Alignment Scores:
Pred. No.: 4.23e-31 Length: 74
Score: 394.00 Matches: 74
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 44.98% Indels: 0
DB: 13 Gaps: 0

US-10-057-510-1 (1-486) x US-10-057-510-2 (1-74)
QY 1 GCACGAGCGTATNAAAATGTTTCCATGTTTATGAAAAGCTGGAGTGTCTGTGT 60
D 1 AlaAargAlaTyr**LysMetPheSerMetPheMetGluLysAlaGlyLysCysTrpCys 20
QY 61 ARAATGCCAGCTCATATAGACTACTCTTCCATGTTTCCCTGCTGCTTAACTGCT 120
D 21 LysMetProLysLeuIleAseThrProPheSerIleValAlaProAlaLeuThrAla 40
QY 121 GTTCTTTCTGCCAGCTTCCTGTTTCCCTGCTGCTGCGGCGACGGCTGTMTCCATGT 180
D 41 ValLeuSerCysGlnLeuAsgCysSerLeuTrpLeuValGlyAlaAargLeu**ProCys 60
QY 181 GCGAAGGTGAAGCATGGACGTGTGGAGGAGCGCTGGAGC 222
D 61 GlyLysValGluGlyMetAspValTrpArgArgTrpSer 74

RESULT 2
US-10-131-410-116
; Sequence 116 Application US/10131410
; Publication No. US20030235915A1
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM TISSUE OF BREAST
; FILE REFERENCE: SCH-1763
; CURRENT APPLICATION NUMBER: US/10131,410
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 09/646,673
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: PCT/DE99/00908
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 116
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-410-116
Alignment Scores:
Pred. No.: 2.66e-30 Length: 247
Score: 387.00 Matches: 94
Percent Similarity: 72.99% Conservative: 6
Best Local Similarity: 68.61% Mismatches: 30
Query Match: 44.18% Indels: 8
DB: 15 Gaps: 2

US-10-057-510-1 (1-486) x US-10-131-410-116 (1-247)
QY 87 TCCTTCTCCATGTTGCCCTGCTCAACTGCTGTTCTTTCGCCAGCTTCGTGTTC 146
D 34 SerPheLeuHisAlaThrHisLeuSer---AlaGluValLysAlaAlaSerLeuPhe 52

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Any X can be any amino acid.
US-10-057-510-2
Alignment Scores:
Pred. No.: 4.23e-31 Length: 74
Score: 394.00 Matches: 74
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 44.98% Indels: 0
DB: 13 Gaps: 0

US-10-057-510-1 (1-486) x US-10-057-510-2 (1-74)
QY 1 GCACGAGCGTATNAAAATGTTTCCATGTTTATGAAAAGCTGGAGTGTCTGTGT 60
D 1 AlaAargAlaTyr**LysMetPheSerMetPheMetGluLysAlaGlyLysCysTrpCys 20
QY 61 ARAATGCCAGCTCATATAGACTACTCTTCCATGTTTCCCTGCTGCTTAACTGCT 120
D 21 LysMetProLysLeuIleAseThrProPheSerIleValAlaProAlaLeuThrAla 40
QY 121 GTTCTTTCTGCCAGCTTCCTGTTTCCCTGCTGCTGCGGCGACGGCTGTMTCCATGT 180
D 41 ValLeuSerCysGlnLeuAsgCysSerLeuTrpLeuValGlyAlaAargLeu**ProCys 60
QY 181 GCGAAGGTGAAGCATGGACGTGTGGAGGAGCGCTGGAGC 222
D 61 GlyLysValGluGlyMetAspValTrpArgArgTrpSer 74

RESULT 3
US-10-264-237-2428
; Sequence 2428 Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL31P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2428
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)_FEATURE
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (11)_FEATURE
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (13)_FEATURE
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (19)_FEATURE
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (187)_FEATURE
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2428
Alignment Scores:
Pred. No.: 8.9e-10 Length: 196
Score: 183.50 Matches: 51
Percent Similarity: 69.14% Conservative: 5
Best Local Similarity: 62.96% Mismatches: 19
Query Match: 20.95% Indels: 6
DB: 15 Gaps: 1
```



```
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 19723
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

Alignment Scores:
Pred. No.: 9 8 Length: 19723
Score: 87.50
Percent Similarity: 33.58%
Best Local Similarity: 30.60%
Query Match: 9.99%
Indels: 43
Gaps: 10
DB: 15

US-10-057-510-1 (1-486) x US-10-084-846A-5 (1-19723)
QY 102 TGCCCTGCTCAACTGCTGT-----TCTTCTCCAGCTTCGTGTTCCTCTGGCT 155
Db 13931 CysAlaGlySerValCysCysAsgThrGlyArgArgProGlyAlaSerPro----- 13948
QY 156 TGTGGGGGCGCGCTGNTCCATGTGCGAAGTGGGAAGGCATGAGCGTGTGGAGGAGCG 215
Db 13949 CysGlyCysSerAlaIleAsgMetTrpSerGlyThrMetSer----- 13962
QY 216 CTGGAGCTGAAGGAATGCGAGCGCTTGGG-----AGAGGCGCAGAGGCTACGCA 266
Db 13963 -----ThrSerThrThrSerProGlySerCysSerArgCysAlaAlaGlySerAla 13979
QY 267 GGGCTGAGTGAAGTGAAGTGGAGCCCTGATNGTCCAGACTCTCAGGACATGCCCCAGT 326
Db 13980 AlaAla-Gly-----CysProProIleThrSerAlaAaGala-----CysProLe 13994
QY 327 CAAGGGTTCGAGCCACNAGNCTTGGCTCATATGCGCATGAAGGGGAACCTGTCATAGAAG 386
Db 13994 u-----ArgAlaAlaCysTrpArgSe 14001
QY 387 C---CCCTCCCT-----GGTTGGTCTCGGCCCTCTGTTCTG 422
Db 14001 rCysProValProAlaAlaProSerArgCysAlaGlySerGlyAspGlyPro----- 14018
QY 423 GAAAACTGGNCTAGNCCCTCGTGTGTTTNGCAAAGCCT 462
Db 14019 -SerThrGlyArgAlaThrGlyTrpMetSerHisGlyPro 14031

RESULT 7
US-09-738-973-41
; Sequence 41, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
```

```
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-41

Alignment Scores:
Pred. No.: 5.99 Length: 163
Score: 86.00
Percent Similarity: 43.22%
Best Local Similarity: 33.05%
Query Match: 9
Indels: 50
Gaps: 5
DB: 5

US-10-057-510-1 (1-486) x US-09-738-973-41 (1-163)
QY 462 AGCTTTGCHAAACACAGGGGNCNTANGANCAGTTTCCAGAACAGAGGGCCAGGACC 403
Db 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluAaGThr 25
QY 402 ACAACAGGAGGGGCTTCTTATGCAAGTTCCTTCATGCCCATATGACCCAGNCCTN 343
Db 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
QY 342 GTGGCTGACCTTGAACCTTGAAGTGGCAATGTC-----TGAGAGTCTGGGACNATCCA 292
Db 46 GlnAlaArgValArgAseLeuAlaGlnSerThrPheProLeuAlaIleGlnGluThrPro 65
QY 291 GGGGCTGCATCTTCTCATCTCAGCCCTCGTAGCCTTCTGCCCTCTCCACAGGGCTCGTCC 232
Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyGlyArgSer 85
QY 231 ATTCTTCAGCTCCAGCGCTCTCCACAGCTCCATGCTTCCACCTTCCACATGGA-- 174
Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101
QY 173 -NACAGCGTCCGCC-----ACAAGCAGAGGGAACACAGAAAGCTGG 132
Db 101 aAlaAlaIleProLeuGluLeuThrGlnSerArgValGlnIleTrp 118

RESULT 8
US-09-854-133-41
; Sequence 41, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-41

Alignment Scores:
Pred. No.: 5.99 Length: 163
Score: 86.00
Percent Similarity: 43.22%
```

```
Best Local Similarity: 33.05% Mismatches: 50
Query Match: 10.24% Indels: 17
DB: 9 Gaps: 5

US-10-057-510-1 (1-486) x US-09-854-133-41 (1-163)

QY 462 AGGCTTTCGNAACCAAGGGGCTTTCATGCAAGTTCCTTCCATGATGAGCCAGGACCC 403
DB 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluArgThr 25
QY 402 ACAACAGGAGGGGGCTTTCATGCAAGTTCCTTCCATGATGAGCCAGGACCC 343
DB 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
QY 342 GTGGCTCGAACCTTCAGTGGCAATGCC-----TGAGACTCTGGGACNATCCA 292
DB 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaGlnGluThrPro 65
QY 291 GGGGCTGCATCTTCCTCAGCCCTGCGTAGCCTTCTGCTCCCTCCAGGGCTCGTCC 232
DB 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85
QY 231 ATTCCTTCAGTCCAGCGCTCTCCACACGTCCTCCATGCTTCCACCTTGCCACATGGA-- 174
DB 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyProSerGlyAl 101
QY 173 -NACAGCGTGGCCCC-----ACAAGCCAGAGGGAACAACGAAGCTGG 132
DB 101 aAlaAlaLysIleProLeuGluLeuThrGlnSerArgValGlnLysIleTrp 118

RESULT 9
US-10-144-649A-41
; Sequence 41, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; PRIOR FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-144-649A-41

Alignment Scores:
Pred. No.: 5.99 Length: 163
Score: 86.00 Matches: 39
Percent Similarity: 43.22% Conservative: 12
Best Local Similarity: 33.05% Mismatches: 50
Query Match: 10.24% Indels: 17
DB: 14 Gaps: 5

US-10-057-510-1 (1-486) x US-10-144-649A-41 (1-163)

QY 462 AGGCTTTCGNAACCAAGGGGCTTTCATGCAAGTTCCTTCCATGATGAGCCAGGACCC 403
DB 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluArgThr 25
QY 402 ACAACAGGAGGGGGCTTTCATGCAAGTTCCTTCCATGATGAGCCAGGACCC 343
DB 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
QY 342 GTGGCTCGAACCTTCAGTGGCAATGCC-----TGAGACTCTGGGACNATCCA 292
DB 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaGlnGluThrPro 65
QY 291 GGGGCTGCATCTTCCTCAGCCCTGCGTAGCCTTCTGCTCCCTCCAGGGCTCGTCC 232
DB 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85
QY 231 ATTCCTTCAGTCCAGCGCTCTCCACACGTCCTCCATGCTTCCACCTTGCCACATGGA-- 174
DB 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyProSerGlyAl 101
QY 173 -NACAGCGTGGCCCC-----ACAAGCCAGAGGGAACAACGAAGCTGG 132
DB 101 aAlaAlaLysIleProLeuGluLeuThrGlnSerArgValGlnLysIleTrp 118

RESULT 10
US-10-156-761-9745
; Sequence 9745, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9745
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9745

Alignment Scores:
Pred. No.: 6.94 Length: 378
Score: 86.00 Matches: 37
Percent Similarity: 35.33% Conservative: 16
Best Local Similarity: 24.87% Mismatches: 37
Query Match: 10.24% Indels: 60
DB: 14 Gaps: 7

US-10-057-510-1 (1-486) x US-10-156-761-9745 (1-378)

QY 342 GTGGCTCGAACCTTCAGTGGCAATGCCATGCTGAGAGTCTGGGACNATCCAGGGGCTGCA 283
DB 5 ValAlaArgGlnLeuGluTrp-----GlyThrVal 14
QY 282 TCTTCATCTCAGCCCTGCGTAGCCTTCTGCTCCCTCCAGGGCTCGTCCATT----- 229
DB 15 SerSerSerThrAlaAlaSerGlyIleAspProIleAlaGluAlaProLeuSerLeu 34
QY 228 -----CCTTCAGCTCCAGCG-----CCTCTCTCCA 205
DB 35 CysAlaArgAlaProHisValProAlaAspArgLeuValAlaGluMetValProProPro 54
QY 204 CACGTCCATGCTTCCACCTTCCACATGAGNACAGCCGTCGCCCCCAAGCCAGAGGGA 145
DB 55 ArgPheAspSerValArgPheAlaThrTyrIle---ProAspProAsnGlnProSerGln 73
QY 144 ACA-----ACGAGCTGGCAAGAAAGACAGCAGT 115
DB 74 ThrGluAlaValArgValLeuGluAspPheAlaThrGlyLeuGlyAlaHisAlaSer 93
QY 114 TAGAGC-----AGGGGCAACAATGGAGAA 91
```

```

Db      94 GlySerGlyArgArgGluPheGlyLysAlaProLysThrProAlaGlyPro 113
QY      90 AGGAGTATCTATTATGAG-----CTT 70
      ||| :||| :|||
Db     114 ArgGlyValTyrLeuAspGlyGlyTyrGlyValGlyLysThrHisLeuAlaSerLeu 133
QY      69 GGGCATTTTACACAGCACATTCCACGCTT 40
      ||| :||| :|||
Db     134 TrpHisAlaThrProAlaGluProGluLeu 143
      ||| :||| :|||

RESULT 11
US-10-425-114-52720
; Sequence 52720, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52720
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700453413_F1r.pap
US-10-425-114-52720

Alignment Scores:
Pred. No.:      8.77      Length:      196
Score:          84.50      Matches:     47
Percent Similarity: 37.57%      Conservative: 21
Best Local Similarity: 25.97%      Mismatches:   63
Query Match:      10.06%      Indels:       50
DB:              12      Gaps:         7

US-10-057-510-1 (1-486) x US-10-425-114-52720 (1-196)

QY      480 GGGANTTTCCTTCTTCTAGGCTTGCNNAAACCAGGGGNCNTANGANCAGTTTTC 421
      ||| :||| :|||
Db     12 GlyTrpTyrPheThrArgAlaCys-----Pro 21

QY      420 GAACAGAGGCCAGGACACACACAGGAGGGGGTCTTATGCAAGTTCCTTCATGC 361
      ||| :||| :|||
Db     22 GluProSerSerThrSerSerThrAlaAlaSerProSerAlaAlaSerAlaPro----- 39

QY      360 CATATGAGCCAAAGNCCTN-----GTGGCTCGAACCTTGTACTGGCCNAATCTCTGAGAG 307
      ||| :||| :|||
Db     40 --LeuSerArgSerLeuTyrAlaValLysProSerSerThrGlyAlaProSerAlaAsp 58
      ||| :||| :|||

QY      306 TCTGGGAGCNATCCAGGGGTGCTATCTTCATCTCAGCCCTG----- 265
      ||| :||| :|||
Db     59 ArgGlySerArgSerGlyThrGlyGlySerValThrAlaLeuAlaAspAlaAla 78

QY      264 -----CGTAGCCTTCTCCCTCCCTCC 244
      ||| :||| :|||
Db     79 AlaPhePheLeuArgAlaProArgPheLeuGlyProThrArgAspLeuIleArgSerThr 98
      ||| :||| :|||

QY      243 CAGGGCTCGTCCATCTCTTCAGCTCCAGCGCTCCTC----- 207
      ||| :||| :|||
Db     99 ArgSerSerLysSerSerSerAlaValAla-ValLeuValAlaAlaGlyAlaGlySerTr 118
      ||| :||| :|||

QY      206 -----CACAGTCCATGCTTCCACCTTGCACATGCGANACAGCCGTGCC 160
      ||| :||| :|||
Db    118 pserAlaAlaAlaValSerSerProGluAlaAlaArgGlyGlyAsnGlySer 138
      ||| :||| :|||

```

```

159  QY  CACAGCCAGAGGGGNAACAACGAGCTGGCAAGAAAGAACAGCAGCAGTTTAGAGCAGGGGCGAAC 100
      :::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
138  Db  rSerArgAlaArgCysArgGlyGluArgAlaValGlyAlaLysGly----- 156
      :::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
199  QY  AATGAGAAAGAGTATCTATTATGAGCTTGGGC---ATTTTACACGACGACTTCCGAC 43
      :::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
157  Db  LeuGluSerGlyValAlaIleAlaProLeuGlyGluCysThrHisGlnHisPheMetAl 176
      :::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
242  QY  C 42
      |
276  Db  a 176

RESULT 12
US-10-425-114-53402
; Sequence 53402, Application US/10425114
; Publication NO. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Steven, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53402
; LENGTH: 250
; TYPE: PRGT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700235435_FLI.pep
; US-10-425-114-53402

```

Alignment Scores:	9.15	Length:	250
Pred. No.:	84.50	Matches:	38
Score:	41.48%	Conservative:	13
Percent Smilarity:	30.83%	Mismatches:	58
Best Local Similarity:	10.06%	Indels:	14
Query Watch:	12	Gaps:	5
DB:			
US-10-057-510-1 (1-486) x US-10-425-114-53402 (1-250)			
QY	425	TTCCAGACAGAGGGCCAGGAGCCAAACAGGAGGGGGCTTCTTATGCAAGTTCCCTT	366
Db	108	PheProArgProGlnAlaProGlnProGlyCysAlaLeu	123
QY	365	CATGCCATATGAGCCAAAGNCTNG	321
Db	124	HisGluIleArgAlaArgAlaargGlnAlaargGlnProHisLeuHisProLeuHisGln	143
QY	320	CAATGT	267
Db	144	ProCysGlyAlaGlyAlaAlaLeuProHisAlaArgLeuHisCysHisLeuAtgPro	163
QY	266	TGGTGAGCTTTCGCCCTCTCCAGGCGCTCGTCATCTTCTATCTCCAGCCC	207
Db	164	ArgGlnGlnGlyProLeuProHisAlaarg	182
QY	206	CACACGTCCATGCTTCCACCTTGCCACATGGANACAGCCGTGCCCCCAACAGCCAGAGG	147
Db	183	HisLeuArgAlaArgGlyArgAlaProHis	201
QY	146	GAACAACGAGCTGGCAAGAAAGACACAGCTAGACGAGGGGCAACAATGGCAAAAGGA	87
Db	202	GlnGluArg-LeuGlyValArgArgHisGlnLuvalseGlyArgGlyArgArgGlnG	221
QY	86	GATCTA	80


```
Db 221 yHisLeu 223
RESULT 13
US-10-425-114-72216
; Sequence 72216, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72216
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3245-492-F2_FLI.ppe
US-10-425-114-72216
Alignment Scores:
Pred. No.: 9.15 Length: 250
Score: 84.50 Matches: 38
Percent Similarity: 41.46% Conservative: 13
Best Local Similarity: 30.88% Mismatches: 58
Query Match: 10.08% Indels: 14
DB: 12 Gaps: 5
US-10-057-510-1 (1-486) x US-10-425-114-72216 (1-250)
QY 425 TTCAGACAGAGGCCAGGACCAACACAGGAGGGGGCTTCTATGCAAGTTCCCTT 366
Db 108 PheProArgGlyProGlnAlaProGlnProGlyCysAlaLeu-----ProVal 123
QY 365 CATGCGATATGAGCAAGNCTNG-----TGCTCGAACCTTGAATGGG 321
Db 124 HisGluIleArgAlaArgAlaArgGlnAlaArgGlnProHisLeuHisProLeuHisGln 143
QY 320 CAATGT-----CCTGAGAGTCTGGGACNATCCAGGGGTCATCTTCATCCTCAGCCC 267
Db 144 ProCysGlyAlaGlyAlaAlaLeuProHisAlaArgLeuHisCysHisLeuArgPro 163
QY 266 TGCGTAGCCTTCGCTCCCTCCAGGGCTGTCATCTTCATCTCCAGCGCTCCTC 207
Db 164 ArgGlnGlnGlnProLeuProHisAlaArg---ValArgAlaLeuProArgGlnVal 182
QY 206 CACAGCTCATGCTTCACCTGCGATGCGATGAGNACAGCCGTCGCCCAACAGCCAGG 147
Db 183 HisLeuArgAlaArgGlyArgAlaProHis---HisHisArgAlaProLeuArgHisGln 201
QY 146 GAACAACAAGTGGCAAGAAAGACAGCAGTACAGGAGGCGCAACAATGGAGAAAGG 87
Db 202 GlnGluArg-LeuGlyValArgArgHisGlnGluValSerGlyArgGlyArgGlnGln 221
QY 86 GTATCTA 80
Db 221 yHisLeu 223
RESULT 14
US-10-425-114-63205
; Sequence 63205, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63205
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73008E02_FLI.ppe
US-10-425-114-63205
Alignment Scores:
Pred. No.: 9.22 Length: 261
Score: 84.50 Matches: 30
Percent Similarity: 38.60% Conservative: 14
Best Local Similarity: 26.32% Mismatches: 33
Query Match: 10.06% Indels: 37
DB: 12 Gaps: 6
US-10-057-510-1 (1-486) x US-10-425-114-63205 (1-261)
QY 417 CAGAGGCCAGGACCAACACAGGAGGGGGCTTCTAT-----GCAAGTTCC 370
Db 43 ArgArgArgGlyThrAlaProArgArgGlyProSerTrpArgGlyProAlaSerGlyAla 62
QY 369 CATTTCATGCGATATGAGCAAGNCTNGTGGTCGACCACTTGAAGTGGGCAATGCTCGA 310
Db 63 ProArgCysSer-----ProTrpSerArgAla 72
QY 309 GAGTCTGGGACNATCCAGGGGCTGCAATCTTCATCTCAGCCCTGCTAGCCTTCTGCC 250
Db 73 ValSerGly-----Thr 76
QY 249 TCTCCAGGCGCTGCTCCATCTTCAGCTCCAGGCGCTCTCCACACGCTCCATGCTTC 190
Db 77 ThrAlaSerGlySerArgAlaProAlaGlyProAlaProTrpPro-----SerSer 93
QY 189 CACTTGGCCACATGGANACAGCGCTGCCCCCAAGCCAGAGGGAAACAACAGAGTGGCA 130
Db 94 ProGluAlaAlaTrpArgArgProSerProProArgProArgGly---HisArgValAla 112
QY 129 AGAAGAACAAGCAGTTAGACGAGGCGCAACAATGGAGAAAGG 88
Db 113 ArgAlaGly-----ThrArgGlyArgGlyAlaArg 123
RESULT 15
US-10-171-311-71
; Sequence 71, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
```

;; PRIOR FILING DATE: 2001-06-13
;; PRIOR APPLICATION NUMBER: US 60/335,936
;; PRIOR FILING DATE: 2001-11-14
;; NUMBER OF SEQ ID NOS: 238
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 71
;; LENGTH: 338
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-171-311-71

Alignment Scores:
Pred. No.: 10.8 Length: 338
Score: 84.00 Matches: 54
Percent Similarity: 37.23% Conservative: 16
Best Local Similarity: 28.72% Mismatches: 56
Query Match: 9.59% Indels: 62
DB: 14 Gaps: 11

US-10-057-510-1 (1-486) x US-10-171-311-71 (1-338)

```
QY 42 GGTGGAGAGTGTGTGTAAATGCCAAGCTCATATAGATACCTCT---TTCTCCAT 98
Db 97 GlyTrpProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerI 113
QY 99 TGTTGCCCTGTCTTAACGTGCTGTTCTTCT-----TGCCAGCT 137
Db 113 eValValProValLeuAlaProGlyLeuGlySerThrArgSerSerAlaLeuCysSerLe 133
QY 138 TCGTTGTTCCCTCTGGCTT----- 156
Db 133 uGlyTyrCysValTyrTyrGlySerAlaPheAlaGlnAlaLeuLeuGlyCysHisAl 153
QY 157 -----GTGGGGGACCGCTGNTCCATGTGGCAAGGTGGAGGCATGGACGTGTGGAGGAG 212
Db 153 aSerLeuGlyHisArgLeuGly---AlaGlyGlnValProGlyLeu-ThrLeuGlyLeu 172
QY 213 GCCTCGAGCTGAAGCA-----ATGCACGAGCCCTGGGAGGAGGCAGAGGC 260
Db 172 hrValGlyIleTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGly 191
QY 261 TAGCAGGG-----CTGAGGATGAAGATGACGCCCTCGATNGTCCC 302
Db 192 AlaSerGlyGlyLeuCysThrLeuIleTyrSerThrGluLeuLysAlaLeuGlnAlaThr 211
QY 303 CAGACTCTCAGGACATTGCCAGTCAAGGTTTCGAGCCACNAGNCTT---GGCTCATAT 359
Db 212 HisThrValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLys 231
QY 360 GGCATGAAGGGGAACCTTGATTAAGAGCCCTCCCTGG-----TTGTGTCCTGG 410
Db 232 GlyLeuLysLysAlaLeuGlyMetGlyProGlyProTyrMetAsnIleLeuTrpAlaTrp 251
QY 411 CCCTCTGTCTGGAACCTGGNTCAGTCCCTGTTTNGCAAGCCTGAGAAAG 470
Db 252 -PheIlePheTrp-----Trp----- 256
QY 471 GGAAGTCCCATGGG 486
Db 257 -----ProHisGly 259
```

Search completed: April 6, 2004, 18:01:08
Job time : 56.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 6, 2004, 17:53:20 ; Search time 19.5 Seconds
(without alignments)

2573.356 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876

Sequence: 1 gcacagcgatnaaaaaat.....aaaggggaantccccatggg 486

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool.p/US10057510/runat_06042004.141609.14770/app_query.fasta_1.647
-DB=Issued Patents AA -QSWT=faetan -SUFFIX=rai -MINMATCH=0_1 -LOOPC1=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10057510 @CGN 1.1.27 @runat_06042004.141609.14770 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp:*
5: /cgn2_6/ptodata/2/iaa/6C.COMB.pcp:*
6: /cgn2_6/ptodata/2/iaa/backfiles.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394	45.0	74	4	US-09-510-252-2
2	95	11.3	195	4	US-09-252-991A-20532
3	87.5	10.4	581	4	US-09-252-991A-20966
4	86	10.2	163	4	US-09-370-838-41
5	86	9.8	446	4	US-09-252-991A-31114
6	82.5	9.8	199	4	US-09-252-991A-25623
7	82.5	9.8	320	4	US-09-252-991A-24634
8	82.5	9.8	447	4	US-09-252-991A-32755
9	82.5	9.4	1652	4	US-09-627-650B-1
10	82.5	9.4	1652	4	US-09-436-063C-1
11	82.5	9.4	1917	4	US-09-627-650B-5
12	82.5	9.4	1917	4	US-09-436-063C-5

13	82.5	9.4	2508	4	US-09-627-650B-7	Sequence 7, Appli
14	82.5	9.4	2508	4	US-09-436-063C-7	Sequence 7, Appli
15	82.5	9.4	2544	4	US-09-627-650B-3	Sequence 3, Appli
16	82.5	9.4	2544	4	US-09-436-063C-3	Sequence 3, Appli
17	82.5	9.4	2601	4	US-09-627-650B-9	Sequence 9, Appli
18	82.5	9.4	2601	4	US-09-436-063C-9	Sequence 9, Appli
19	82	9.8	127	4	US-09-252-991A-36902	Sequence 26902, A
20	82	9.8	150	4	US-09-252-991A-31728	Sequence 31728, A
21	80.5	9.6	206	4	US-09-252-991A-34085	Sequence 24085, A
22	80	9.5	204	4	US-09-252-991A-36183	Sequence 26183, A
23	79.5	9.1	299	3	US-09-188-930-192	Sequence 192, App
24	79.5	9.5	339	4	US-09-252-991A-18253	Sequence 18253, A
25	79	9.0	147	4	US-09-252-991A-32565	Sequence 32565, A
26	79	9.0	176	4	US-09-252-991A-35290	Sequence 25290, A
27	79	9.4	229	4	US-09-252-991A-36254	Sequence 26254, A
28	78.5	9.3	202	4	US-09-252-991A-32054	Sequence 32054, A
29	78.5	9.0	299	3	US-09-188-930-332	Sequence 332, App
30	78.5	9.0	299	4	US-09-312-283C-192	Sequence 192, App
31	78.5	9.0	299	4	US-09-312-283C-332	Sequence 332, App
32	78.5	9.3	347	4	US-09-252-991A-19498	Sequence 19498, A
33	77.5	9.2	144	4	US-09-252-991A-31237	Sequence 31237, A
34	77.5	8.8	340	4	US-09-252-991A-19460	Sequence 19460, A
35	77.5	9.2	487	4	US-09-252-991A-29392	Sequence 29392, A
36	77.5	8.8	519	4	US-09-252-991A-19734	Sequence 19734, A
37	77.5	9.2	527	4	US-09-370-838-216	Sequence 216, App
38	77.5	9.2	777	4	US-09-252-991A-20378	Sequence 20378, A
39	77	8.8	168	4	US-09-252-991A-32502	Sequence 32502, A
40	77	9.2	422	4	US-09-252-991A-30314	Sequence 30314, A
41	77	9.2	447	4	US-09-252-991A-25916	Sequence 25916, A
42	76.5	8.7	360	4	US-09-252-991A-32452	Sequence 32452, A
43	76.5	9.1	1729	4	US-09-553-690-2	Sequence 2, Appli
44	76	9.0	136	4	US-09-252-991A-24892	Sequence 24892, A
45	76	9.0	148	4	US-09-252-991A-28538	Sequence 28538, A

ALIGNMENTS

RESULT 1
US-09-510-252-2
; Sequence 2, Application US/09510252
; Patent No. 6372490
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Yang Meijia
; APPLICANT: Schulz, Vincent
; APPLICANT: Curagen Corporation
; TITLE OF INVENTION: MDM INTERACTING PROTEIN AND METHODS OF USE THEREOF
; FILE REFERENCE: 15966-524 MDM US
; CURRENT APPLICATION NUMBER: US/09/510,252
; CURRENT FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: USSN 60/121,192
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: USSN 60/122,643
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Any X can be any amino acid.
US-09-510-252-2

Alignment Scores: 2.26e-36 Length: 74
Pred. No.: 394.00 Matches: 74
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 44.98% Gaps: 0
DB: 4

US-10-057-510-1 (1-486) x US-09-510-252-2 (1-74)

QY	1	GCACGAGCGGTATNAAAAAATGTTTCCCATGTTTATGGAAAAGCGTGGGAAGTGCTGGTGT	60
Db	1	AlaArgAlaTyr***LysMetPheSerMetPheMetGluLysAlaGlyLysCysTyrCys	20
QY	61	AAAAATGCCCAAGCTCATAATAGATACTCCTTTCCCATTTGCCCTGCTCTAACTGT	120
Db	21	LysMetProLysLeuIleAlaAspThrProPheSerIleValAlaProAlaLeuThrAla	40
QY	121	GTTCTTTCTTGGCCAGCTTCGTGTGTTCCCTCTGGCTTGTGGGGCAGCGCTGTTNCCATGT	180
Db	41	ValLeuSerCysGlnLeuArgCysSerLeuTrpLeuValGlyAlaArgIleu***ProCys	60
QY	181	GGCAGGTGGAAAGGCATGACGTGTGGAGGAGCGCTGGAGC	222
Db	61	GlyLysValGluGluYMetAspValTrpArgArgTrpSer	74

RESULT 2

```

US-09-252-991A-20532
; Sequence 20532, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20532
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20532

```

US-09-370-838-41
; Sequence 41, Application US/09370838
; Patent No. 644425
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:
Pred. No.: 0.225 Length: 163
Score: 86.00 Matches: 39
Percent Similarity: 43.22% Conservative: 12
Best Local Similarity: 33.05% Mismatches: 50
Query Match: 10.24% Indels: 17
DB: 4 Gaps: 5

US-10-057-510-1 (1-486) x US-09-370-838-41 (1-163)
; SEQ ID NO 41
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapien

US-09-370-838-41
QY 462 AGGCTTTCGAAACACGAGGGGNCNTANGNCAGTTTCGAGAACAGAGGGCCAGGACC 403
Db 11 ArgLeuSerLeuArgPro-----SerileProGluLeuSerGluArgThr 25
QY 402 ACAACGAGGGGGGCTTCTATGCAAGTTCCCTTCATCCCATATGAGCCAGAGNCCTN 343
Db 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
QY 342 GTGGCTCGAACCTTACTGGGCAATGTCC-----TGAGAGTCTGGGGACNATCCA 292
Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65
QY 291 GGGGTGATCTTATCTTCAGCCCTGCTAGCCTTCGCTCCCTCCAGGGGCTCTCC 232
Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85
QY 231 ATTCTTCAGCTCCAGCGCTCTCCACACGCTCCATGCTTCCACCTTGCCACATGGA-- 174
Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101
QY 173 -NACAGCGTGGCCCTC-----ACAAGCCAGGGGAAACAAGAGCTGG 132
Db 101 aAlaAlaIlysileProLeuGluLeuThrGlnSerArgValGlnIlysileTrp 118

RESULT 5
US-09-252-991A-31114
; Sequence 31114, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

US-09-252-991A-31114
; SEQ ID NO 31114
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:
Pred. No.: 0.6 Length: 199
Score: 82.50 Matches: 41
Percent Similarity: 40.91% Conservative: 13
Best Local Similarity: 31.06% Mismatches: 58
Query Match: 9.82% Indels: 20
DB: 4 Gaps: 6

US-10-057-510-1 (1-486) x US-09-252-991A-31114 (1-446)
; SEQ ID NO 31114
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:
Pred. No.: 0.316 Length: 446
Score: 86.00 Matches: 43
Percent Similarity: 34.56% Conservative: 4
Best Local Similarity: 31.62% Mismatches: 37
Query Match: 9.82% Indels: 52
DB: 4 Gaps: 9

US-10-057-510-1 (1-486) x US-09-252-991A-31114 (1-446)
; SEQ ID NO 31114
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:
Pred. No.: 0.6 Length: 199
Score: 82.50 Matches: 41
Percent Similarity: 40.91% Conservative: 13
Best Local Similarity: 31.06% Mismatches: 58
Query Match: 9.82% Indels: 20
DB: 4 Gaps: 6

US-09-252-991A-31114
; SEQ ID NO 31114
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:
Pred. No.: 0.316 Length: 446
Score: 86.00 Matches: 43
Percent Similarity: 34.56% Conservative: 4
Best Local Similarity: 31.62% Mismatches: 37
Query Match: 9.82% Indels: 52
DB: 4 Gaps: 9

US-10-057-510-1 (1-486) x US-09-252-991A-31114 (1-446)
; SEQ ID NO 31114
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

QY 29 TGTATTGGAAGGCTGGGAAGTGTGCTGTAATAATGCCAAGCTCATAATAGATATC 88
Db 65 CysAlaTrpIysArgArgCysSerAlaGlyAla-GlyArg----- 77
QY 89 CTTTCTCCATTGTGCTTAACTGCTGTTCTTTTCCAGCTTCGTTGCC 148
Db 78 -----CysArgCysSerSer-----SerGlyThrProAlaGlyGlyHisPr 91
QY 149 TCTGGCTTGTGGGGCAGCGCTGTNTCCATGTGGCAGGTGGAAGCATGGACGTGGA 208
Db 91 oProAlaSerGlyArgGluAlaArgGly---TrpPro----- 102
QY 209 GGAGGGCTGGAGCTGAAGGAATGGACGAGCCCTGGGAGGAGG-----GCAGAAGS 259
Db 103 -----ArgArgTrpProArgProGlyArgArgSerSerAlaAlaG 117
QY 260 CTAGGAGGGCTGAGG-----ATGAAGATGACAGCCCTCGATNGTCCCA 304
Db 117 yArgCysGlyCysArgGluCysSerGlyAlaGly-ArgCysGluProProValGlyPro- 136
QY 305 GACTCTCAGACATTCGCCAGTCAAGGTTTCGAGCCACNAGNCTTGGCTCATATGGCAT 364
Db 137 -----AlaGlyHisAlaProValArg----- 143
QY 365 GAAGGGGAATTCGCATA---AGAAGCCCTCTCTGTTGTGGT 405
Db 144 --ProGlyArgCysAlaGlyArgAspArgLeuProGlyCysGly 157

RESULT 6
US-09-252-991A-25623
; Sequence 25623, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

US-09-252-991A-25623
; SEQ ID NO 25623
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:
Pred. No.: 0.6 Length: 199
Score: 82.50 Matches: 41
Percent Similarity: 40.91% Conservative: 13
Best Local Similarity: 31.06% Mismatches: 58
Query Match: 9.82% Indels: 20
DB: 4 Gaps: 6

US-09-252-991A-25623
; SEQ ID NO 25623
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

Alignment Scores:
Pred. No.: 0.6 Length: 199
Score: 82.50 Matches: 41
Percent Similarity: 40.91% Conservative: 13
Best Local Similarity: 31.06% Mismatches: 58
Query Match: 9.82% Indels: 20
DB: 4 Gaps: 6

US-10-057-510-1 (1-486) x US-09-252-991A-25623 (1-199)
QY 462 AGCTTTGNAACACAGGGGNCCTPANGANCAGTTTCCAGAACAGAGGCCACAGACC 403
Db 40 ArgThrCysAla---ProAlaGlyGlnArgAlaGlnProGlyGlnArgAspArgPro 58
QY 402 ACACACAGGAGGGG-----GCTCTTATGCAAGTTCCCTTCATGCGCATATG 355
Db 59 ValGlnArgProGlyGlyArgThrAlaLeuArgArgArgThrGlySerArgHisLeu 78
QY 354 AGCCAAAGNCTGTGCTCGAACCTTGTCTGAGCAATGCTGAGAGTCTGGGACNAT 295
Db 79 AlaGlyAlaAlaCysGlyArgAlaLeuGlnArgAlaLeuSerGluGlyAlaGlyGluPhe 98
QY 294 CCAGGG-----GCTGCATCTTCATCTCCAGCCCTG 265
Db 99 ProGlyArgThrLeuAlaGlyAlaGluProProAlaGlyArgGlyValArgValLeu 118
QY 264 CGTAGCCTTCTGCTCCCTCCAGGCGC---TCGTCCATTCCTTCA---GCTCCAGCGCT 211
Db 119 ArgArgHisProAlaPheArgGlyArgProAlaIleProAspLeuAlaProAlaSer 138
QY 210 CTTCCACAGTCATGCTTCCACCTTGCACATGGANACAGCGTGGCCCCCAGGACA 151
Db 139 Pro-----AlaLeuAlaPhe-LeuLeuProArgGlyAlaSerAlaGlyArgSerArgG 156
QY 150 GAGGGAACACAGAGCTGGCAAGAAAGAACAGCA 117
Db 156 uArgPheArgArgGlyAlaValGlnLeuProAla 167
RESULT 7
US-09-252-991A-24634
; Sequence 24634, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24634
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24634
Alignment Scores:
Pred. No.: 0.703 Length: 320
Score: 82.50 Matches: 38
Percent Similarity: 38.62% Conservative: 18
Best Local Similarity: 26.21% Mismatches: 69
Query Match: 9.82% Indels: 20
DB: 4 Gaps: 6
US-10-057-510-1 (1-486) x US-09-252-991A-24634 (1-320)
QY 456 TGCNAAAACAGGGGNCCTPANGANCAGTTTCCAGAACAGAGGCCACCAACC 397
Db 12 CysArgArgProSerAla---SerAlaAlaGlyProArgThrProProArgThrSerPhe 30
QY 396 AGGAGGGGGCTTCTTATGCAAGTTCCTTCATGCGCATATGAGCAAGNCTGTGGCT 337
Db 31 SerAlaSerArgProArgValThrSerProThrCysGlyAspCysAlaArgSerThrSer 50
QY 336 GGAACCTTACTGGGCAATGCTCTGAGAGTCTGGGACNATCCA-----GGG 289

51 Arg-----TrpArgThrSerSerProAlaProSerProCysAlaSerAlaSer 67
288 GCTGCATCTTCATCTCCAGCCCTGCTAGCCTTCTGCTCCCTCCAGGCGCTGCTCATT 229
Db 68 AlaArgSerGlySerArgAlaCysAlaSerThrAlaSerCysSerProIle 87
228 CTTTCA-----GCTCCAGCGCTCTCCACACGTCATGCTCCACCTTCCACCTTGC 181
Db 88 SerSerProAlaCysArgAlaThrAlaProPro-ArgArgSerSerProGlyThrAlaPr 107
180 ACATGGANACAGCGTGGCCCCCAGAGCCAGAGGGAACACGAAGC----- 135
Db 107 aArgAlaCysThrAlaAlaGlyAlaThrSerThrAlaIleArgArgSerAlaCysGlyMetPr 127
134 -TGCAAGAAAGAACAGCAGTGTAGCAGGCGGCAACAATGGAGAAAGAGTATCTATTAT 76
Db 127 oTrp-----LysSerAlaMetArgAlaSerSerThrArgArgSerSerCysAla 145
75 GAGCTTGGGCATT 53
145 aAsnThrGlyLeu 149
RESULT 8
US-09-252-991A-32755
; Sequence 32755, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32755
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32755
Alignment Scores:
Pred. No.: 0.787 Length: 447
Score: 82.50 Matches: 40
Percent Similarity: 37.98% Conservative: 9
Best Local Similarity: 31.01% Mismatches: 36
Query Match: 9.82% Indels: 44
DB: 4 Gaps: 7
US-10-057-510-1 (1-486) x US-09-252-991A-32755 (1-447)
QY 401 CAACAGGAGGGGGCTTCTTATGCAAGTTCCTTCATGCTCCATATGAGCCAGNCTNG 342
Db 16 ArgProAlaArgGly-----ProLeuHis-----AlaAlaArgGln 27
341 TGGCTC-----GAACCTTGTACTGGCAATGCTCCTGAGAGTCTGGGACN 297
Db 28 TrpLeuProGlyLeuArgArgAlaProLeuAlaGlyHisArgProAlaAlaProGlyPro 47
296 ATCCAGGGGCTGCATCTTCAT----- 276
48 ProProHisAlaArgArgHisArgGlySerAlaProArgArgGlySerArgThrAlaAla 67
275 -----CCTCAGCCCTCGGTAGCC-----TTCTGCGCTCCT 246
68 SerArgProArgProAlaAlaAlaArgArgProAlaAlaProGlyTrpSerAlaProGly 87
245 CCCAGGGCTCGCTCCATTCCTTTCAGTCCAGCGCC---TCCTCCACAGTCCATGCTTCC 189
Db 88 ProArgAlaAlaProLeuProValArg-AlaGlyAsnAlaValArgProCysArgPr 107

QY 188 ACCTGTCCACATGGANACAGCGGTGCTCCCAAGCCAGGAGGAGCAACAGAGCTGGCAA 129
Db 107 oPro-AlaalaTrpProAlaProCysAlaHis-AlaIleAGAGHisArg----- 123
QY 128 GAAGAAGCAGAGTGTAGAGCA 108
Db 124 ---ArgSerAlaAlaArgAla 129

RESULT 9

US-09-627-650B-1

; Sequence 1, Application US/09627650B

; Patent No. 6406872

; GENERAL INFORMATION:

; APPLICANT: Bamber, Bruce

; APPLICANT: Jorgensen, Erik

; TITLE OF INVENTION: Nematoe Neuromuscular Junction GABA Receptors and

; TITLE OF INVENTION: Methods Related Thereto

; FILE REFERENCE: 21101.0009U3

; CURRENT APPLICATION NUMBER: US/09/627,650B

; PRIOR FILING DATE: 2000-07-28

; PRIOR APPLICATION NUMBER: 09/436,063

; PRIOR FILING DATE: 1999-11-08

; PRIOR APPLICATION NUMBER: 60/107,727

; PRIOR FILING DATE: 1998-11-09

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1652

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-09-627-650B-1

Alignment Scores:

Pred. No.: 1.22 Length: 1652
Score: 82.50 Matches: 37
Percent Similarity: 35.71% Conservative: 8
Best Local Similarity: 29.37% Mismatches: 40
Query Match: 9.42% Indels: 41
DB: 4 Gaps: 6

US-10-057-510-1 (1-486) x US-09-627-650B-1 (1-1652)

QY 99 TGTGTCCTGCTCTAAC-----TGCTGTCTTCTTCTTCCAGCTTCGTTG 143
Db 524 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla----- 537
QY 144 TTCCTCTGCTGTGGGGCAGCGCTGNTCCATGTGCGAAGGTGGAGGCATGGAGCT 203
Db 538 -----CysAlaCysAlaThrAlaAlaCysThr-CysGlyThrThr-CysCysThrThrC 555
QY 204 GT-----GGAGGAGGCGCTGGAGCTGAAGGAATGGACGAGCCCT 242
Db 555 ysglyThrAlaThrCysGlyAlaGlyGlyThrGlyAlaThrGly--GlyAlaAlaCy 574
QY 243 GGGAGGAGGCGAGAGGTGACGAGGCTGAGGATGAGATGAGAGCCCTCATATGCC 302
Db 574 sglyGlyThrThrThrAlaThr-----AlaCysThr----- 584
QY 303 CAGACTCTCAGGACATTCGCCAGTCAAGGTTTCGAGCCACNAGGCTTGGCTCATATGCC 362
Db 585 -----AlaGlyThrCysAlaAlaAlaGlyAlaThrThrAlaAlaCysAla----- 599
QY 363 ATGAAGGGGAACCTGCATAGAAGCCCTCCCTCGTTGTGCTGCTGCTGCTGCTGCTG 422
Db 600 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 617
QY 423 GAAGAACTGGN 432
Db 617 aalathrGly 620

RESULT 10

US-09-436-063C-1

; Sequence 1, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematoe Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1652
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-1

Alignment Scores:
Pred. No.: 1.22 Length: 1652
Score: 82.50 Matches: 37
Percent Similarity: 35.71% Conservative: 8
Best Local Similarity: 29.37% Mismatches: 40
Query Match: 9.42% Indels: 41
DB: 4 Gaps: 6

US-10-057-510-1 (1-486) x US-09-436-063C-1 (1-1652)

QY 99 TGTGTCCTGCTCTAAC-----TGCTGTCTTCTTCTTCCAGCTTCGTTG 143
Db 524 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla----- 537
QY 144 TTCCTCTGCTGTGGGGCAGCGCTGNTCCATGTGCGAAGGTGGAGGCATGGAGCT 203
Db 538 -----CysAlaCysAlaThrAlaAlaCysThr-CysGlyThrThr-CysCysThrThrC 555
QY 204 GT-----GGAGGAGGCGCTGGAGCTGAAGGAATGGACGAGCCCT 242
Db 555 ysglyThrAlaThrCysGlyAlaGlyGlyThrGlyAlaThrGly--GlyAlaAlaCy 574
QY 243 GGGAGGAGGCGAGAGGTGACGAGGCTGAGGATGAGATGAGAGCCCTCATATGCC 302
Db 574 sglyGlyThrThrThrAlaThr-----AlaCysThr----- 584
QY 303 CAGACTCTCAGGACATTCGCCAGTCAAGGTTTCGAGCCACNAGGCTTGGCTCATATGCC 362
Db 585 -----AlaGlyThrCysAlaAlaAlaGlyAlaThrThrAlaAlaCysAla----- 599
QY 363 ATGAAGGGGAACCTGCATAGAAGCCCTCCCTCGTTGTGCTGCTGCTGCTGCTGCTG 422
Db 600 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 617
QY 423 GAAGAACTGGN 432
Db 617 aalathrGly 620

RESULT 11

US-09-627-650B-5

; Sequence 5, Application US/09627650B

; Patent No. 6406872

; GENERAL INFORMATION:

; APPLICANT: Bamber, Bruce

; APPLICANT: Jorgensen, Erik

; TITLE OF INVENTION: Nematoe Neuromuscular Junction GABA Receptors and

; TITLE OF INVENTION: Methods Related Thereto

; FILE REFERENCE: 21101.0009U3

; CURRENT APPLICATION NUMBER: US/09/627,650B

; PRIOR FILING DATE: 2000-07-28

; PRIOR APPLICATION NUMBER: 09/436,063

; PRIOR FILING DATE: 1999-11-08

; PRIOR APPLICATION NUMBER: 60/107,727


```

; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1917
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-5

Alignment Scores:
Pred. No.: 1.28 Length: 1917
Score: 82.50 Matches: 37
Percent Similarity: 35.71% Conservatives: 8
Best Local Similarity: 29.37% Mismatches: 40
Query Match: 9.42% Indels: 41
DB: 4 Gaps: 6

US-10-057-510-1 (1-486) x US-09-627-650B-5 (1-1917)
QY 99 TGTGCCCCCTGCTCTAAC-----TGCTGTCTCTTCTTCCAGCTTCGTTG 143
Db 703 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla-----716
QY 144 TTCCTCTGCTGTGGGGCAGCGCTGNTCCATGTGGCAAGTGGAGGCGATGGAGCT 203
Db 717 -----CysAlaCysAlaThrAlaAlaCysThr-CysGlyThrThr-CysCysThrThrC 734
QY 204 GT-----GGAGGAGCGCTGGAGCTGAAGGAATGGAGCGCCCT 242
Db 734 ySGlyThrAlaThrCysGlyAlaGlyGlyThrGlyAlaThrGly--GlyAlaAlaCy 753
QY 243 GGGAGGAGGCGAGAGCTACGAGGCTGAGGATGAAGTGCAGCCCTCGATNGTCCC 302
Db 753 sGlyGlyThrThrAlaThr-----AlaCysThr-----763
QY 303 CAGACTCTCAGGACATTGCCAGTCAAGGGTTCAGGCCACNAGNCTTGGCTCATATGGC 362
Db 764 -----AlaGlyThrCysAlaAlaAlaGlyAlaThrThrAlaAlaCysAla-----778
QY 363 ATGAAGGGAACTGCTAATAAGAGCCCTCCCTGCTGTGTGCTGCTGCTGCTGCTGCTG 422
Db 779 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 796
QY 423 GAAACTGCGN 432
Db 796 aAlaThrGly 799

RESULT 13
US-09-627-650B-7
; Sequence 7, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-7

Alignment Scores:
Pred. No.: 1.4 Length: 2508
Score: 82.50 Matches: 37
Percent Similarity: 35.71% Conservatives: 8
Best Local Similarity: 29.37% Mismatches: 40
Query Match: 9.42% Indels: 41
DB: 4 Gaps: 6

US-10-057-510-1 (1-486) x US-09-627-650B-7 (1-2508)
QY 99 TGTGCCCCCTGCTCTAAC-----TGCTGTCTCTTCTTCCAGCTTCGTTG 143
Db 543 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla-----556
QY 144 TTCCTCTGCTGTGGGGCAGCGCTGNTCCATGTGGCAAGTGGAGGCGATGGAGCT 203
```

Db 557 -----CysAlaCysAlaThrAlaAlaCysThr-CysGlyThrThr-CysCysThrThrC 574
QY 204 GT-----GGAGGAGCGCTGGAGCTGAAGGAATGAGCGCCCT 242
Db 574 ySgLYThrAlaThrCysGlyAlaGlyGlyThrGlyAlaThrGly--GlyAlaAlaCy 593
QY 243 GGGAGGAGGCGAGAGGCTAGCAGGCTGAGGATGAAGATGAGCCCTGGATNGTCCC 302
Db 593 sGlyGlyThrThrAlaThr-----AlaCysThr----- 603
QY 303 CAGACTCTCAGACATGCCAGTCAAGGCTGAGCCACNAGNCTTGGCTCATATGCG 362
Db 604 -----AlaGlyThrCysAlaAlaGlyAlaThrThrAlaAlaCysAla----- 618
QY 363 ATGAAGGGGAATTCATAGAGAGCCCTCCCTGGTGTGGTCTGCTGCTGTTCTG 422
Db 619 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 636
QY 423 GAAAACTGGN 432
Db 636 aAlaThrGly 639
RESULT 14
US-09-436-063C-7
; Sequence 7, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-7
Alignment Scores:
Pred. No.: 1.4 Length: 2508
Score: 82.50 Matches: 37
Percent Similarity: 35.71% Conservative: 8
Best Local Similarity: 29.37% Mismatches: 40
Query Match: 9.42% Indels: 41
DB: 4 Gaps: 6
US-10-057-510-1 (1-486) x US-09-436-063C-7 (1-2508)
QY 99 TGTGCGCCCTGCTCTAAC-----TGCTGTCTTTCTTCTGCCAGCTTCGTTG 143
Db 543 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla----- 556
QY 144 TTCCTCTGCTGTGGGCGACGGCTGTTCATGTGCAAGGTGGAAGGATGAGCGT 203
Db 557 -----CysAlaCysAlaThrAlaAlaCysThr-CysGlyThrThr-CysCysThrThrC 574
QY 204 GT-----GGAGGAGGCGCTGAGCTCAAGGAATGAGCGCCCT 242
Db 574 ySgLYThrAlaThrCysGlyAlaGlyGlyThrGlyAlaThrGly--GlyAlaAlaCy 593
QY 243 GGGAGGAGGCGAGAGGCTAGCAGGCTGAGGATGAAGATGAGCCCTGGATNGTCCC 302
Db 593 sGlyGlyThrThrAlaThr-----AlaCysThr----- 603
QY 303 CAGACTCTCAGACATGCCAGTCAAGGCTGAGCCACNAGNCTTGGCTCATATGCG 362
Db 604 -----AlaGlyThrCysAlaAlaGlyAlaThrThrAlaAlaCysAla----- 618
QY 363 ATGAAGGGGAATTCATAGAGAGCCCTCCCTGGTGTGGTCTGCTGCTGTTCTG 422
Db 619 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 636
QY 423 GAAAACTGGN 432
Db 636 aAlaThrGly 639

Db 604 -----AlaGlyThrCysAlaAlaGlyAlaThrThrAlaAlaCysAla----- 618
QY 363 ATGAAGGGGAATTCATAGAGAGCCCTCCCTGGTGTGGTCTGCTGCTGTTCTG 422
Db 619 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 636
QY 423 GAAAACTGGN 432
Db 636 aAlaThrGly 639
RESULT 15
US-09-627-650B-3
; Sequence 3, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-3
Alignment Scores:
Pred. No.: 1.41 Length: 2544
Score: 82.50 Matches: 37
Percent Similarity: 35.71% Conservative: 8
Best Local Similarity: 29.37% Mismatches: 40
Query Match: 9.42% Indels: 41
DB: 4 Gaps: 6
US-10-057-510-1 (1-486) x US-09-627-650B-3 (1-2544)
QY 99 TGTGCGCCCTGCTCTAAC-----TGCTGTCTTTCTTCTGCCAGCTTCGTTG 143
Db 543 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla----- 556
QY 144 TTCCTCTGCTGTGGGCGACGGCTGTTCATGTGCAAGGTGGAAGGATGAGCGT 203
Db 557 -----CysAlaCysAlaThrAlaAlaCysThr-CysGlyThrThr-CysCysThrThrC 574
QY 204 GT-----GGAGGAGGCGCTGAGCTCAAGGAATGAGCGCCCT 242
Db 574 ySgLYThrAlaThrCysGlyAlaGlyGlyThrGlyAlaThrGly--GlyAlaAlaCy 593
QY 243 GGGAGGAGGCGAGAGGCTAGCAGGCTGAGGATGAAGATGAGCCCTGGATNGTCCC 302
Db 593 sGlyGlyThrThrAlaThr-----AlaCysThr----- 603
QY 303 CAGACTCTCAGACATGCCAGTCAAGGCTGAGCCACNAGNCTTGGCTCATATGCG 362
Db 604 -----AlaGlyThrCysAlaAlaGlyAlaThrThrAlaAlaCysAla----- 618
QY 363 ATGAAGGGGAATTCATAGAGAGCCCTCCCTGGTGTGGTCTGCTGCTGTTCTG 422
Db 619 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 636
QY 423 GAAAACTGGN 432
Db 636 aAlaThrGly 639

Wed Apr 7 07:32:31 2004

Search completed: April 6, 2004, 17:59:23
Job time : 25.5 secs

us-10-057-510-1.ra1

Page 8

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 6, 2004, 17:48:14 ; Search time 54.5 Seconds
(without alignments)
5039.196 Million cell updates/sec

Title: US-10-057-510-1
Perfect score: 876
Sequence: 1 gcacgaggtatnaaaat.....aaagggaantccccctggg 486

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cn2.1/USPTO spool p/US10057510/runat_06042004_141608_14698/app_query.fasta_1.647
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -CUTFM=ptc -NORF=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10057510 @CGN_1.1.81/runat_06042004_141608_14698 -NCFU=6 -ICFU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394	45.0	74	3	Aab08845
2	387	44.2	247	2	Aay48595 Human bre
3	183.5	20.9	196	5	Abb90052 Human pol
C	93.5	11.1	575	4	Abg21970 Novel hum
5	92.5	10.6	288	4	Abg04148 Novel hum
C	91	10.8	1373	4	Abb62368 Drosophil
C	89.5	10.7	647	2	Aaw04327 Rat petr
C	89.5	10.7	1708	5	Abp69375 Human pol
9	89	10.2	239	5	Abg60084 Human DIT
C	88.5	10.5	247	6	Aae35942 Lolium pe

C	11	86	10.2	163	2	AAY29526	Human lun
C	12	86	10.2	163	3	AAB44416	Human lun
C	13	86	10.2	163	4	AAE13757	Human lun
C	14	86	10.2	163	7	ADD66349	Human lun
C	15	86	10.2	163	7	AD87603	Human blo
	16	84	9.6	338	2	AD873955	Duffy blo
	17	84	9.6	338	2	AAW54857	Novel sapi
	18	84	9.6	338	5	ABB76196	Human Duf
	19	84	9.6	338	6	ABP81813	Human cer
	20	84	9.6	338	6	ABR2081	Human cer
	21	83.5	9.9	266	4	ABR70975	Drosophil
	22	83.5	9.9	266	4	ABR53281	Protein s
	23	83.5	9.9	3502	4	ABR53282	Drosophil
	24	83	9.9	202	4	AAW79674	Human pro
	25	83	9.9	213	4	AAW78690	Human pro
	26	83	9.9	221	4	AAW78690	Human pro
	27	83	9.9	239	3	AAW53294	Human col
	28	83	9.9	239	4	AAW5034	Human col
	29	83	9.9	426	4	AAW62628	Chlorella
	30	83	9.5	542	4	ABR11016	Human sec
	31	83	9.5	2087	4	AAW31516	Amino aci
	32	83	9.5	2091	3	AAW12000	Rat p3103
	33	82.5	9.4	213	4	AAW19618	Human dia
	34	82.5	9.8	511	5	ABG93255	C. albica
	35	82.5	9.4	1652	6	ADA15715	C. elegan
	36	82.5	9.4	1917	6	ADA15719	C. elegan
	37	82.5	9.4	2508	6	ADA15721	C. elegan
	38	82.5	9.4	2544	6	ADA15717	C. elegan
	39	82.5	9.4	2601	6	ADA15723	C. elegan
	40	82	9.8	363	4	ABR68590	Drosophil
	41	82	9.8	1386	7	ADE59291	Rat Prote
	42	82	9.8	1386	7	ADE59295	Rat Prote
	43	82	9.8	19938	6	ABP76682	Streptomy
	44	81.5	9.7	580	4	ABG21040	Novel hum
	45	81.5	9.3	963	4	ABG02396	Novel hum

ALIGNMENTS

RESULT 1

ABR08845

ID AAE08845 standard; protein; 74 AA.

AC AAE08845;

DT 02-JAN-2001 (first entry)

XX A human MDM2 interacting polypeptide (MDMIP).

XX Human; MDM2 interacting polypeptide; MDMIP; MDM2; cell cycle progression; cell differentiation; cancer; glioma; squamous cell carcinoma; breast cancer; astrocytoma; leukemia; lymphoma; tumorigenesis; gene therapy.

OS Homo sapiens.

XX Key

FH Key Location/Qualifiers

FT Misc-difference 5 /note= "unknown amino acid encoded by NAA"

FT Misc-difference 58 /note= "unknown amino acid encoded by TNT"

XX WO2000050590-A1.

XX 31-AUG-2000.

XX 23-FEB-2000; 2000WO-US004582.

XX 23-FEB-1999; 99US-0121192P.

XX 03-MAR-1999; 99US-0122643P.

XX 22-FEB-2000; 2000US-00510252.

XX (CURA-) CURAGEN CORP.

XX PI Nandabalan K, Yang M, Schulz VP;
 XX DR WPI: 2000-558398/51.
 XX DR N-PSDB; AA75041.
 XX FT Novel MDM2 interacting protein useful for treating or preventing
 XX FT disorders involving aberrant levels of MDM2 and/or MDMV-interacting
 XX FT proteins, comprises a specific amino acid sequence.
 XX PS Claim 10; Fig 1; 78pp; English.
 XX CC The present sequence represents a human MDM2 interacting polypeptide
 CC (MDMIP). MDMIP was identified using a yeast two hybrid system, using a
 CC fragment of MDM2 as the bait protein. The MDMIP polypeptide is useful for
 CC detecting and removing MDM2 polypeptides in a biological sample by
 CC forming MDM2-MDMIP complexes. MDMIP and MDM2 are useful to identify
 CC compounds or other agents which modulate the activity of MDM2 and/or
 CC MDMIP-mediated processes. Agents that modulate the function of MDMIP/MDM2
 CC complexes are useful for treating and preventing a disease or disorder
 CC involving aberrant levels of MDM2 or MDMIP. MDMIP is also useful for
 CC treating diseases caused by aberrant levels of expression of MDM2 genes,
 CC such as disorders of cell cycle progression, cell differentiation, and
 CC transcriptional control, including cancers such as human sarcoma, glioma,
 CC squamous cell carcinoma, breast cancer, astrocytoma, leukemia and
 CC lymphoma, and tumorigenesis. MDMIP and MDM2 nucleic acids are useful in
 CC gene therapy
 XX SQ Sequence 74 AA;

Alignment Scores:
 Pred. No.: 2.01e-32 Length: 74
 Score: 394.00 Matches: 74
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 44.98% Indels: 0
 DB: 3 Gaps: 0

US-10-057-510-1 (1-486) x AAB08845 (1-74)

QY 1 GCACGAGCGTATNAAAAATTTTCCATGTTTATGAAAGCGTGGAGTCTGGTGT 60
 DB 1 AlaArgAlaTyr***LysMetPheSerMetPheMetGluLysAlaGlyLysCysTrpCys 20
 QY 61 AAATGCGCCAGCTCATATATAGACTCCCTTCTCCATTTCTGCCCTGCTCTAATGCT 120
 DB 21 LysMetProLysLeuIleAlaAspThrProPheSerIleValAlaProAlaLeuThrAla 40
 QY 121 GTTCTTTCTGCCAGCTTCGTTGTTCCCTCTGCTTGTGGGGCAGCGGTGNTCCATGT 180
 DB 41 ValLeuSerCysGlnLeuArgCysSerLeuTrpLeuValGlyAlaArgLeu***ProCys 60
 QY 181 GGCAGCTGGAGCATGACGCTGTGGAGGAGCGCTGGAGC 222
 DB 61 GlyLysValGluGlyMetAspValTrpArgArgTrpSer 74

RESULT 2
 AAY48585
 ID AAY48585 standard; protein; 247 AA.

XX AC AAY48585;

DI 08-DEC-1999 (first entry)

XX DE Human breast tumour-associated protein 46.
 XX KW Expressed sequence tag; EST; human; breast; cancer; gene therapy;
 XX KW treatment; tumour; cytostatic; medicament.

XX OS Homo sapiens.

XX FN DE19813839-A1.

XX

PD 23-SEP-1999.

XX 20-MAR-1998; 98DE-01013839.

XX 20-MAR-1998; 98DE-01013839.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosentahl A;

XX WPI: 1999-528981/45.

XX N-PSDB; AA233655.

XX Human nucleic acid sequences and protein products from tumor breast
 XX tissue, useful for breast cancer therapy.

XX Claim 22; 163; 188pp; German.

XX This invention describes novel human nucleic acid sequences from tumor
 CC breast tissue which have cytostatic activity. The nucleic acid sequences
 CC can be used to produce and isolate full-length gene sequences. They can
 CC be used to express proteins, which can be used as tools to find an
 CC activity against breast cancer. The sequences can be used in sense or
 CC antisense form. They are especially useful for medicaments for gene
 CC therapy to treat breast cancer. AAY48540-Y48617 represent protein
 CC fragments encoded by the expressed sequence tags described in the method
 CC of the invention

XX SQ Sequence 247 AA;

Alignment Scores:
 Pred. No.: 1.4e-31 Length: 247
 Score: 387.00 Matches: 94
 Percent Similarity: 72.99% Conservative: 6
 Best Local Similarity: 68.61% Mismatches: 30
 Query Match: 44.18% Indels: 8
 DB: 2 Gaps: 2

US-10-057-510-1 (1-486) x AAY48585 (1-247)

QY 87 TCCTTTCTCCATGTTGGCCCTGCTCTAACTGCTGTTCTTTCTTCCAGCTTCGTTGTTTC 146
 DB 34 SerPheLeuHisAlaThrHisHisLeuSer---AlaGluValIysAlaAlaSerLeuPhe 52
 QY 147 CCTCTGCTGTTGGGGCAGCGCTGNTTC-CATGTGGCAAGGTGAAGCATGGACGTGT 205
 DB 53 ProLeuAlaCysGlyGlyThrAlaValLeuHisValAlaAargTrpLysAlaTrpThrCys 72
 QY 206 GGAGGAGCGCTGAGCTGAAGGAATGGACGAGCCCTGGAGGAGGAGGAGGAGGAGGAG 265
 DB 73 GlyGlyGlyAlaGlyAlaGlyAlaGlyMetAspGluProTrpGluGluGlyArgArgLeuArg 92
 QY 266 AGGCTGAGGATGAAGTGAAGCCCTGATGATGTCCTCCAGACTCTCAGGACATTCGCCAG 325
 DB 93 ArgAlaGluAspGluAspAlaAlaProGlyTrpSer-GlnThrLeuArgThrCysProAl 112
 QY 326 TCAAGGGTTCCGAGCCACNAGNCTTGGCTCATATGGCATGAAGGGGAGGAGGAGGAGGAG 385
 DB 112 aGlnGlyLeuArgAlaThrGlyLeuAlaSerTyrglyMetArgGlySerTrpHis-ArgS 132
 QY 386 GCCCCTCCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 445
 DB 132 erProLeuPro-AlaValValLeu-ProSerValLeuGlnThr---AlaLeuSerProLe 150
 QY 446 GGTTTTNGCAAGCCTGAAGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 486
 DB 150 uAlaLeuCysGlnAlaTrpArg-ArgAlaValProHisGly 163

RESULT 3
 ABB90052
 ID ABB90052 standard; protein; 196 AA.
 XX AC ABB90052;

XX		24-MAY-2002 (first entry)	
DT		Human polypeptide SEQ ID NO 2428.	
XX		Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KW		antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;	
KW		vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KW		cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;	
KW		neurological disease; infection; human; secreted protein.	
XX		Homo sapiens.	
OS		WO200190304-A2..	
PV		29-NOV-2001.	
XX		18-MAY-2001; 2001WO-US016450.	
PF		19-MAY-2000; 2000US-0205515P.	
XX		(HUMA-) HUMAN GENOME SCI INC.	
PA		Birse CE, Rosen CA;	
PI		WPI; 2002-122018/16.	
PT		N-PSDB; ABL90461.	
DR		Novel 1405 isolated polypeptides, useful for diagnosis, treatment and	
XX		prevention of neural, immune system, muscular, reproductive,	
PT		Gastrointestinal, pulmonary, cardiovascular, renal and proliferative	
PT		disorders.	
XX		Claim 11; SEQ ID NO 2428; 2081pp + Sequence Listing; English.	
XX		The invention relates to novel Genes (ABL89449-ABL90853) and proteins	
FS		(ABB99040-ABB90444) useful for preventing, treating or ameliorating	
CC		medical conditions e.g. by protein or gene therapy. The genes are	
CC		isolated from a range of human tissues disclosed in the specification.	
CC		The nucleic acids, proteins, antibodies and antagonists are useful in	
CC		the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and	
CC		ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,	
CC		cancer, gastrointestinal tract, liver, lung, or urogenital; (b) immune	
CC		disorders e.g. Addison's disease, allergies, autoimmune haemolytic	
CC		anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,	
CC		multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)	
CC		cardiovascular disorders such as myocardial ischaemias; (d) wound healing	
CC		; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)	
CC		infectious diseases such as viral, bacterial, fungal and parasitic	
CC		infections. Note: The sequence data for this patent did not form part of	
CC		the printed specification, but was obtained in electronic format directly	
CC		from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX		Sequence 196 AA;	
SQ			
Alignment Scores:			
Pred. No.:	2,51e-10	Length:	196
Score:	183.50	Matches:	51
Percent Similarity:	69.14%	Conservative:	5
Best Local Similarity:	62.96%	Mismatches:	19
Query Match:	20.95%	Indels:	6
DB:	5	Gaps:	1
US-10-057-510-1 (1-486) x AB990052 (1-196)			
QY	254	AGAAGGCTACGCAGGGCTGAGGATCAAGATCAGCCCTCGATTNGTCCCAGACTCTCAG	313
Db	38	ArgArgLeuArgArgAlaGluAspGlnAspAlaProGlyTrpSer-GlnThrLeuAr	57
QY	314	GACATTGCCAGTCGAAGGTTCGACCACNAGNCCTGGCTCATATGGCATGAGGGGA	373
Db	57	gthrCySProAlaGInGlyLeuArgAlaThrGlyLeuAlaserTyroGlyMetArgGlyse	77

Qy 374 CTTGCATAGAAGCCCTCCCTCGTGTGTGGTCTCTGGCCCTCTGTCTTGGAACCTGGT 43

Db 77 rTrpHis-ArgSerProLeuPro-AlaValValleu-ProSerValleuGlnThr---Al 95

Qy 434 CNTAGNCCCTCGTGTGTTCGCAAGCTGAGAAAGGGAANNTCCCATGGG 486

Db 95 aleuSerProLeuAlaLeuCysGlnAlaTrpArg-ArgAlaValProHisGly 112

RESULT 4

ID ABG21970

XX ABG21970 standard; protein; 575 AA.

XX AC ABG21970;

XX AC

XX DT 18-FEB-2002 (first entry)

XX DT

XX DE Novel human diagnostic protein #21961.

XX DE

XX XN Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX XN food supplement; medical imaging; diagnostic; genetic disorder.

XX XN

XX OS Homo sapiens.

XX OS

XX WO200175067-A2.

XX PN

XX PN 11-OCT-2001.

XX PD

XX PF 30-MAR-2001; 2001WO-US008631.

XX PF

XX PR 31-MAR-2000; 2000US-00540217.

XX PR

XX PR 23-AUG-2000; 2000US-00649167.

XX PR

XX PA (HYSE-) HYSEQ INC.

XX PA

XX PI Drmanac RT, Liu C, Tang YT;

XX PI

XX DR WPI; 2001-639362/73.

XX DR

XX DR N-PSDB; AAS6157.

XX DR

XX PT New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity.

XX PT

XX PS Claim 20; SEQ ID NO 52329; 103pp; English.

XX PS

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX CC sequences. (II) is useful as hybridisation probes, polymerase chain

XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX CC and in recombinant production of (II). The polynucleotides are also used

XX CC in diagnostics as expressed sequence tags for identifying expressed

XX CC genes. (I) is useful in gene therapy techniques to restore normal

XX CC activity of (II) or to treat disease states involving (II). (II) is

XX CC useful for generating antibodies against it, detecting or quantitating a

XX CC polypeptide in tissue, as molecular weight markers and as a food

XX CC supplement. (II) and its binding partners are useful in medical imaging

XX CC of sites expressing (II). (I) and (II) are useful for treating disorders

XX CC involving aberrant protein expression or biological activity. The

XX CC polypeptide and polynucleotide sequences have applications in

XX CC diagnostics, forensics, gene mapping, identification of mutations

XX CC responsible for genetic disorders or other traits to assess biodiversity

XX CC and to produce other types of data and products dependent on DNA and

XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

XX CC amino acid sequences of the invention. Note: The sequence data for this

XX CC patent did not appear in the printed specification, but was obtained in

XX CC electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX CC

SQ Sequence 575 AA;

Alignment Scores:

Pred. No.: 0.809 Length: 575

Score: 93.50 Matches: 36

Alignment Scores:		
Pred. No.:	0.809	Length:
score:	93.50	Matches:
		36
		575

Percent Similarity: 40.32% Conservative: 14
Best Local Similarity: 29.03% Mismatches: 57
Query Match: 11.13% Indels: 17
DB: 4 Gaps: 3

US-10-057-510-1 (1-486) x ABG21970 (1-575)

QY 447 CCAGGGGNCNTANGANCAGTTTCCAGAACAGAGGGCCAGG---ACCACACAGGGAG 391
DB 198 ProGlyHisGlnArgSerLeuMetProProSerArg***ArgAlaThrSerThrAlaTrp 217
QY 390 GGGGCTTCTTATGCAAGTCCCTTCATGCCATATGAGCCAGGCTGTGGCTCGAAC 331
DB 218 ArgAlaAa-gthAlaSerThrProArgCys-----LeuGlyProArgPro 232
QY 330 CTGTGCTGGGCAATCTCGAGAGTCTGGGACNATCCAGGGGCTGCATCTCTCTCA 271
DB 233 AlaProTrpSerCys-----AlaSerLeuSerSerSer 243
QY 270 GCCTGCGTAGCTTCTGCTCTCCAGGGCTGCTCCATCTCTTCCAGCTCCAGCGCT 211
DB 244 HisArgProAsnThrGlnProSerProSerProThrSerValArgArgAlaArgLeuAla 263
QY 210 CTCCACACGTCCTCCCTTCCACCTTGCACATGANNACAGCCGCTGCCCCCAAGCCA 151
DB 264 ThrProSerSerHisGlyTrpAlaLeuAlaProTrpThrProMetAlaProThrGlnPro 283
QY 150 GAGGACACAGCAAGTGCAGAGAGACAGACAGTTCAGACAGGGGGCAACATGGAGAA 91
DB 284 ThrThrTrpThrArg-TrpGlyLysThrValSerGlyArg-GlnMetAsnThrTrpArgA 303
QY 90 AGGAGTAT 83
DB 303 rgProTrp 305

RESULT 5
ID ABG04148
ABG04148 standard; protein; 288 AA.

XX AC ABG04148;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #4139.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KM food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.

XX WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS68335.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.

XX Claim 20; SEQ ID NO 34507; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: the sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 288 AA;

Alignment Scores:
Pred. No.: 0.887 Length: 288
Score: 92.50 Matches: 37
Percent Similarity: 37.67% Conservative: 18
Best Local Similarity: 25.34% Mismatches: 50
Query Match: 10.56% Indels: 41
DB: 7 Gaps:

US-10-057-510-1 (1-486) x ABG04148 (1-288)

QY 40 AAGCTGGGAAGTGTGCTGTAATAATGCCAAGCTCATATAGATACTCTTCTTCATT 99
DB 48 LysSerHisGluLeuPheCysLysMetGluProGlnValLeuGluProArgAlaVal 67
QY 100 GTT-----GCCCTGCTCTAACTGCTGCTGCTTCTTCTGCGAGCTTCTTCTCCCTC 150
DB 68 LeuGlnAspGlyAlaProGlyProGlyArgAlaThrSerCysSerAlaArg----- 84
QY 151 TGGCTTGTGGGGCAGCGCTGTNTCCATGTGGCAAGTGGCAAGGCATGACGCTGTGGAGG 210
DB 85 -----LysGlyArgGlyProGluLysProValGlnGlyLeuPro----- 97
QY 211 AGCGCTGGAGCTGAAGGAATGGA-----CGAGCCCTGGGAGGAGGAGAGGCTACG 264
DB 98 -----AsnGlySerValArgAlaHisSerGlyGlyArgAlaPro 111
QY 265 CAGGCTGAGGATGAGATGCAGCCCTGGATNGTCCCGAGACTCTCAGACATTCGCCA 324
DB 112 Gln-----ProSerProArgGlyHisGlyPro 120
QY 325 GTCAAGGGTTGAGCCACNAGNCTTGGCTCATATGCGCATGAAGGGGAACCTGCATAAGA 384
DB 121 GlyArgGly***ThrAlaAlaProLeuProHisLeu-----CysProLeu 135
QY 385 AGCCCTCTCCCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 444
DB 136 ThrProValLeuLeuGlnGly***GlyProAsp***TrpProLeuGly----- 151
QY 445 TGGTTTTTNGCAAGCT 462
DB 152 TrpAlaThrMetArgPro 157

RESULT 6
ID ABB62368
ABB62368 standard; protein; 1373 AA.
XX ABB62368;
XX DT 26-MAR-2002 (first entry)


```

XX Drosophila melanogaster polypeptide SEQ ID NO 13896.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS Drosophila melanogaster.
FN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL06471.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 13896; 21bp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (AB857737-
CC AB872072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1373 AA;

Alignment Scores:
Pred. No.: 1.78 Length: 1373
Score: 91.00 Matches: 46
Percent Similarity: 41.40% Conservative: 19
Best Local Similarity: 29.30% Mismatches: 63
Query Match: 10.83% Indels: 29
DB: 4 Gaps: 7

US-10-057-510-1 (1-486) x ABB62368 (1-1373)
QY 441 GGNCTANGANCAGTTTTCAGAACAGAGGCGCAGGACCAACACAGGAGGGGCTTCT 382
Db 1029 GlyValValAenThrAsnProValArgProArgSerAlaSerSerHisGlyHisSer 1048
QY 361 TATGCAAGT-----TCCCTTCATGCCATATGAGCCAGNCTNGTGCT 337
Db 1049 ValGlySerThrSerAlaProThrSerProSerValHisAlaSerSerGlnValIleLys 1068
QY 336 CGAACCTTGAGTGGGCAATGCTCTGAGAGTCTGGGAC-----NATCCAGGG 289
Db 1069 ArgSerSerProAlaArgSerGlnAlaSerGlnGlyAlaLeuAspLeuThrProArg 1088
QY 288 GTGCA---TCCTTCATCCAGCTGGTAGCCCTTCGCTCTCCAGGGGCTGCTCC 232
Db 1089 AlaAlaProThrSerSerSerSerArgSerProLeuProLysGluLysProValSer 1108
QY 231 ATTCTTTCAGTCCAGCGC-----CTCTCCCA 205
Db 1109 ProProSerLeuPro-ArgSerProSerGlySerSerHisAlaSerAlaAsnIleLeuTh 1128

```

```

QY 204 CACGTCCATGCTTCCACC-----TTGCCACATGGANACAGCCGTGCCCC 160
Db 1128 rSerProLeuProProThrValGlyLeuAspCysLeuProProGlyLeuGlnHis---Hi 1147
QY 159 CACAGCCAGAGGGAACACAGAGCTGGCAGAGAACAGACAGCTAGAGCAGGCGGCAAC 100
Db 1147 sLeuGlnGlnHisGlnHisLeuMetGlnGlnGlnAlaAlaValAlaAlaAlaAla 1167
QY 99 AATGGAGAAAGAGTATCTATTATGAGCTTGGGCACTTTTACACCAACAC 51
Db 1167 aalaglnHisHis---HisHisGlnGlnMetAlaAlaLeuHisGlnHis 1182

RESULT 7
AAW04327
ID AAW04327 standard; protein; 647 AA.
XX
AC AAW04327;
XX
DT 16-JAN-1997 (first entry)
XX
DE Rat petrin.
XX
KW Petrin; neurite outgrowth associated protein; CNS;
KW central nervous system; myelin; protein phosphatase 2C; stroke;
KW neurodegeneration.
XX
OS Rattus sp.
XX
PN WO9632476-A1.
XX
PD 17-OCT-1996.
XX
PF 12-APR-1996; 96WO-CA000214.
XX
PR 13-APR-1995; 95US-00421701.
XX
PA (MOUN) MOUNT SINAI HOSPITAL CORP.
XX
PI Roach A, Lozano A, Labes M, Roder J;
XX
DR WPI; 1996-477127/47.
DR N-PSDB; AAT38486.
XX
PT Assay for substance that modulates response of neuronal cells - and
PT neurite growth associated protein, Petrin, useful in conditions involving
PT nerve damage resulting from traumatic injury, stroke or CNS degenerative
PT disorders.
XX
PS Claim 10; Fig 24; 119pp; English.
XX
CC Rat petrin (AAW04327) is a protein involved in modulating neurite growth
CC inhibition. Its amino sequence was deduced from a cDNA sequence
CC (AAT38486) derived from an adult rat brain cDNA library. Petrin is a new
CC member of the protein phosphatase 2C family, and is expressed in neurons
CC in brain tissue, partic. in the Purkinje cells of the cerebellum. Petrin,
CC and antibodies raised against it, can be used to modulate neurite growth
CC and axonal regeneration
XX
SQ Sequence 647 AA;

Alignment Scores:
Pred. No.: 2.17 Length: 647
Score: 89.50 Matches: 39
Percent Similarity: 41.60% Conservative: 13
Best Local Similarity: 31.20% Mismatches: 52
Query Match: 10.65% Indels: 21
DB: 2 Gaps: 6

US-10-057-510-1 (1-486) x AAW04327 (1-647)
QY 447 CCAGGGGNGNCTANGANCAGTTTTCAGAACAG-----AGGCCAGGACCAACACAGGAG 391
Db 17 ProGlyGlyLeuSerSerAlaHisAlaGluThrGlyArgGlyAlaGlyLysArgPro 36

```


KW gastrointestinal disorder; transport disorder; Gene therapy; kidney;
KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;
KW thymus.

OS Homo sapiens.

FN WO200220754-A2.

XX 14-MAR-2002.

XX 29-AUG-2001; 2001WO-US027127.

XX 05-SEP-2000; 2000US-02297472.

XX 05-SEP-2000; 2000US-02297482.

XX 05-SEP-2000; 2000US-02297492.

XX 05-SEP-2000; 2000US-02297502.

XX 05-SEP-2000; 2000US-02297512.

XX 05-SEP-2000; 2000US-02305832.

XX 06-SEP-2000; 2000US-02305032.

XX 06-SEP-2000; 2000US-02305142.

XX 06-SEP-2000; 2000US-02305152.

XX 06-SEP-2000; 2000US-02305172.

XX 06-SEP-2000; 2000US-02305182.

XX 06-SEP-2000; 2000US-02305192.

XX 06-SEP-2000; 2000US-02305932.

XX 06-SEP-2000; 2000US-02305972.

XX 06-SEP-2000; 2000US-02305982.

XX 06-SEP-2000; 2000US-02305992.

XX 06-SEP-2000; 2000US-02306102.

XX 06-SEP-2000; 2000US-02308652.

XX 06-SEP-2000; 2000US-02308882.

XX 07-SEP-2000; 2000US-02309512.

XX 07-SEP-2000; 2000US-02311632.

XX 07-SEP-2000; 2000US-02311672.

XX (INCY-) INCYTE GENOMICS INC.

XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;

XX Jones AU, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;

XX Miyama MC, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;

XX Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;

XX Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;

XX WPI: 2002-383054/41.

XX N-PSDB; ABK71875.

XX An isolated polynucleotide useful in diagnostics and therapeutics.

XX Claim 29; Page 614-615; 686pp; English.

XX The invention relates to human diagnostic and therapeutic (dithp)

XX polynucleotides and their associated polypeptides (DITHP polypeptides).

XX The sequences of the invention are used in the treatment and diagnosis of

cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers

(e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,

cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or

thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,

psoriasis, osteoporosis), viral infections, bacterial infections, fungal

infections, parasitic infections, developmental disorders (e.g. anaemia,

epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),

endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders

(e.g. obesity, diabetes), neurological disorders (e.g. stroke,

amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal

disorders (e.g. ulcerative colitis, lysinuria) and transport disorders

(e.g. myotonic dystrophy, catatonias, peripheral neuropathy). Sequences

ABG59943-ABG60220 represent human DITHP polypeptides of the invention

XX SQ Sequence 239 AA;

Alignment Scores:

Pred. No.: 1.98 Length: 239

Score: 89.00 Matches: 38

Percent Similarity: 31.85% Conservative: 5

Best Local Similarity: 28.15% Mismatches: 48

Query Match: 10.16% Indels: 44

DB: 5 Gaps: 7

US-10-057-510-1 (1-486) x ABG60084 (1-239)

QY 132 CCAGCTTCGTTGTTCCCT-----CTGGCTTGTGGGGCAGC 167

DB 18 ProAlaGlyLeuLeuProArgAlaAlaLeuGlnA-gvalHisLeuSerCysGlyGly*** 37

QY 168 GCTGTTTCATGTGGCAAGGTGAAGGCATGACGCTGTGGAGGCGCGCTGAGCTGAAG 227

DB 38 -----GluGluLeuSerGlyGlyArgPheGly-----AlaSerAsnLeuGluProAla 53

QY 228 GAATGACGAGCCCTGGGAGGAGGCA----- 254

DB 54 ThrTIPValAspProGlyArg***Leu***SerGlyCysGlyAlaProArgGluLeu 73

QY 255 -----GAAGCTACGACGGCTGAGGATGAAGATG 284

DB 74 GlyProProArgSerSerHisSerAlaAspGlyLeuArgGlyArgGlnGlyThr 93

QY 285 CAGCCCTCGATNGTCCCGAGACTCTCAGGACATTGCC-----AGTCAAGGGTTC 335

DB 94 GlyValIleGlnThrAlaTrpAlaThrHisTIPLeuProCysHisThrAlaGlyPhe 113

QY 336 GAGCCACNA---GGNCTTGGCTCATATGCGCATGAAGGGAACCTGCATGAAGAGCCCTC 392

DB 114 SerProSerLeuGlySerGlySerValAlaProSerGlyProLeuLeuAlaSerProLeu 133

QY 393 CCCTGG-----TTGTGCTCCTGGCCCTCT 416

DB 134 AlaTyrGlyGlyArgLeuLeuSerSerTIPSerArgProSer 148

RESULT 10

AAE35942

ID AAE35942 standard; protein; 247 AA.

XX AC AAE35942;

XX DT 17-JUN-2003 (first entry)

XX DE Lolium perenne defensin f (LpDBFF) protein.

XX KW Ryegrass; thionin; TH; thaumatin-like protein; TL; defensin; DEF; ER;

XX KW elicitor-responsive protein; disease resistance; plant defence response;

XX KW protein storage; pest resistance; genetic marker; gene therapy; antipest;

XX KW agricultural; LpDBFF.

XX OS Lolium perenne.

XX PH Key Location/Qualifiers

FT Misc-difference 243

FT /note= "Encoded by NTC, Xaa corresponds to any amino acid"

XX WO200288359-A1.

XX PD 07-NOV-2002.

XX PF 01-MAY-2002; 2002WO-AU000539.

XX PR 02-MAY-2001; 2001AU-00004735.

XX PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX PA (AGRE-) AGRESEARCH LTD.

XX PI Spangenberg G, Sawbridge TI, Ong EK, Emmerling M;

XX DR WPI; 2003-201227/19.

XX DR N-PSDB; AAD54405.

XX PT New ryegrass or fescue nucleic acid encoding a thionin, thaumatin-like,

PT elicitor-responsive or defensin polypeptide, for modifying disease and/or
 PT plant resistance, plant defence response and/or protein storage in a
 PT plant.

XX Claim 19; Fig 33; 195pp; English.

XX The invention relates to novel ryegrass (Lolium) or fescue (Festuca)
 CC nucleic acids encoding thionin (TH), thaumatin-like (TL), elicitor-
 CC responsive (ER) or defensin (DEF) proteins and polynucleotides encoding
 CC such proteins. Sequences of the invention are useful to modify disease
 CC and/or pest resistance, plant defence response and/or protein storage in
 CC plants. The nucleotide sequence information and/or single nucleotide
 CC polymorphisms of the nucleic acid are useful as genetic markers. They are
 CC also useful in gene therapy. The present sequence is ryegrass DBFF
 CC (UpDEff) protein

XX Sequence 247 AA;

Alignment Scores:
 Pred. No.: 2.25 Length: 247
 Score: 88.50 Matches: 39
 Percent Similarity: 38.46% Conservative: 6
 Best Local Similarity: 33.33% Mismatches: 41
 Query Match: 10.54% Indels: 31
 DB: 6 Gaps: 7

US-10-057-510-1 (1-486) x AAE35942 (1-247)

QY 341 TGGCTCGAACCTTG-----ACTGGCAATGCTCTGAGAGTCTGGGACNATC 294
 Db 83 TrpLeuAlaLeuAlaPheAlaIleThrAsnArgGlyIleGlnSerPheSerAspHis 102
 QY 293 CAGGGGCTGATCTTCATCTCAGCCCTGGTAGCCTTC-----TGCCCTCTCCAGG 240
 Db 103 ProLeuLeuThrLeuHisProGlnArgLeuValProPheIleGluIleProProAsn 122
 QY 239 GCT-----CGTCCATTCTTCAGCTCCAGCGCTCTCCACACGCTCCATGCT 192
 Db 123 AlaGlyHisAlaTyrLeuAspThrLeuProSerProArgLeuLeuAlaThrHisLeuPro 142
 QY 191 ---TCCACCTTGGCACATGGANACAGC-----CGT 165
 Db 143 LeuSerMetLeuProHisGlyMetSerSerAlaGlyCysArgIleValTyrLeuCysArg 162
 QY 164 GCGCCCAAGCCAGCAGCGGACACACGAGCTGCAAGAAAGACACACAGCTTAGACAGG 105
 Db 163 AspProLysAspAlaLeuValSerArg---TriHisPheGluAsnLysIleArgLysGly 181
 QY 104 GCAACAATGGAG-----AAAGGAGTATCT 81
 Db 182 PheAsnMetGluLeuGluLysAlaPheAlaMetPheCysGluGlyValSer 198

RESULT 11
 AAY29526
 ID AAY29526 standard; protein; 163 AA.

XX AC AAY29526;

XX 13-OCT-1999 (first entry)

XX Human lung tumour protein LT86-10 predicted amino acid sequence.

XX Human, lung tumour protein; therapy; diagnosis; lung cancer; vaccine;
 XX immunotherapy; detection; inhibition.

XX Homo sapiens.

XX WO9938973-A2.

XX 05-AUG-1999.

XX 26-JAN-1999; 99WO-US001642.

XX

PR 28-JAN-1998; 98US-00015022.
 PR 28-JAN-1998; 98US-00015029.
 PR 18-MAR-1998; 98US-00040828.
 PR 18-MAR-1998; 98US-00040831.
 PR 23-JUL-1998; 98US-00121191.
 PR 23-JUL-1998; 98US-00121192.
 PR 22-DEC-1998; 98US-00219245.

XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Frudakis TN, Mohamath R;

XX WPI; 1999-479187/40.

DR N-ESDB; AAZ07203.

XX Lung tumor specific polynucleotides for inhibiting the development of
 PT lung cancer.

XX Example 2; Page 77-78; 171pp; English.

XX The present invention describes lung tumour specific polynucleotides and
 CC tumour antigens. AAZ07144 to AAZ07246 and AAZ08301 to AAZ08325 represent
 CC specifically claimed polynucleotides, and AAY29486 to AAY29571 represent
 CC amino acid sequences from the present invention. The lung tumour specific
 CC polynucleotides and polypeptides can be used in pharmaceutical
 CC compositions and vaccines to inhibit the development of lung cancer. They
 CC can also be used to detect lung cancer in a patient. Probes and
 CC antibodies derived from the lung tumour sequences are useful in detection
 CC of lung cancer

XX Sequence 163 AA;

Alignment Scores:
 Pred. No.: 3.75 Length: 163
 Score: 86.00 Matches: 39
 Percent Similarity: 43.22% Conservative: 12
 Best Local Similarity: 33.05% Mismatches: 50
 Query Match: 10.24% Indels: 17
 DB: 2 Gaps: 5

US-10-057-510-1 (1-486) x AAY29526 (1-163)

QY 462 AGGCTTTGCNNAACACAGGGGNGCTANGANCAGTTTTCAGAACAGAGGGCCAGACC 403
 Db 11 ArgLeuSerLeuArgPro-----SerileProGluLeuSerGluArgThr 25
 QY 402 ACAACAGGAGGGGGCTCTTATGCAAGTCCCTTCATGCCATATGAGCCAGNCTN 343
 Db 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
 QY 342 GTGGCTCGAACCTTGACTGGGCAATGCTCC-----TGAGAGTCTGGGACNATCCA 292
 Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65
 QY 291 GGGGCTGCATCTTCATCTCAGCCCTGGTAGCCTTCTGCGCTCTCCAGGGCTCGTCC 232
 Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyGlyArgSer 85
 QY 231 ATTCTCTCAGCTCCAGGCGCTCTCCACAGCTCCATGCTTCCACCTTGCACATGGA-- 174
 Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101
 QY 173 -NACAGCGTCCGCC-----ACAGCCAGAGGGAACACGAGCTGG 132
 Db 101 aAlaAlaLysIleProLeuGluLeuThrGlnSerArgValGlnLysIleTrp 118

RESULT 12

AAB44416

ID AAB44416 standard; protein; 163 AA.

XX AAB44416;

XX 05-FEB-2001 (first entry)

XX Human lung tumour-specific antigen encoded by cDNA #26.

XX Lung tumour protein; lung cancer; cytostatic; vaccine.

XX Homo sapiens.

XX W0200060077-A2.

XX 12-OCT-2000.

XX 30-MAR-2000; 2000WO-US008560.

XX 02-APR-1999; 99US-00285323.

XX 09-AUG-1999; 99US-00370838.

XX 30-DEC-1999; 99US-00476235.

XX 03-MAR-2000; 2000US-00519809.

XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Mohamath R, Secrist H;

XX WPI; 2000-638466/61.

XX N-PSDB; AAC79071.

XX Novel lung tumor polypeptides and polynucleotides, useful for detecting,

XX monitoring or treating cancer, especially lung cancer.

XX Claim 1; Page 102-103; 243pp; English.

XX The present sequence is given in a specification relating to compounds

XX for therapy and diagnosis of lung cancer. Polypeptides comprising at

XX least an immunogenic part of a lung tumour protein are disclosed. The

XX polypeptides are useful for inhibiting the development of cancer,

XX especially lung cancer. Samples of T cells expressing the polypeptides

XX may be used to inhibit the development of cancer. The polypeptides are

XX also useful for detecting and monitoring the progression of cancer,

XX especially lung cancer

XX Sequence 163 AA;

XX Alignment Scores:

XX Pred. No.: 3.75 Length: 163

XX Score: 86.00 Matches: 39

XX Percent Similarity: 43.22% Conservative: 12

XX Best Local Similarity: 33.05% Mismatches: 50

XX Query Match: 10.24% Indels: 17

XX DB: 3 Gaps: 5

XX US-10-057-510-1 (1-486) x AAB44416 (1-163)

XX QY 462 AGGCTTTCGNAACACACAGGGGNGTANGANCAGTTTCCAGACAGAGCCAGGACC 403

XX Db 11 ArgLeuSerLeuArgPro-----SerileProGluLeuSerGluArgThr 25

XX QY 402 ACAACAGGAGGGGGCTCTTATGCAAGTTCCTCATGCTATGATGAGCCAGGAGGACC 343

XX Db 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerProAla 45

XX QY 342 GTGGTCGACCCCTTGACTGGGCAATGCC-----TGAGAGTCGGGGACNATCCA 292

XX Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65

XX QY 291 GGGGCTGATCTTCATCTCAGCCCTGCTAGCCCTTCCTCCCTCCAGGGCTCGTCC 232

XX Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85

XX QY 231 ATTCTTCAGCTCCAGCGCTCTCCACACGTCCTCATGCTCCACCTTGCCACATGGA-- 174

XX Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101

XX QY 173 -NACAGCCGTGCCCC-----ACAAGCCAGAGGGAACAACGAGCTGS 132

Db 101 aAlaAlaLysileProLeuGluLeuThrGlnSerArgValGlnLysileTrp 118

RESULT 13

AAE13757

ID AAE13757 standard; protein; 163 AA.

XX AAE13757;

XX 26-FEB-2002 (first entry)

XX Human lung tumour-specific protein L786-10.

XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;

XX antisense-therapy; vaccine; immune response; lung cancer; L786-10.

XX Homo sapiens.

XX W0200172295-A2.

XX 04-OCT-2001.

XX 28-MAR-2001; 2001WO-US009991.

XX 29-MAR-2000; 2000US-00538037.

XX 05-JUN-2000; 2000US-00588937.

XX 18-AUG-2000; 2000US-00640878.

XX 22-SEP-2000; 2000US-0234517P.

XX 01-NOV-2000; 2000US-00704512.

XX 14-DEC-2000; 2000US-00738973.

XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;

XX Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;

XX WPI; 2001-639201/73.

XX N-PSDB; AAD23146.

XX New human lung-specific polynucleotides and polypeptides for the

XX diagnosis and treatment of disease e.g. lung cancer.

XX Example 2; Page 158-159; 378pp; English.

XX The invention relates to isolated lung tumour-specific proteins and their

XX corresponding cDNA molecules. Lung tumour-specific proteins and their

XX antigen-presenting cells are useful for stimulating and/or expanding T

XX cells specific for a tumour protein, and for inhibiting the development

XX of cancer. The invention also relates to a composition useful for

XX stimulating an immune response, and for treating cancer. The lung tumour

XX specific oligonucleotide is useful in gene therapy and for diagnosis,

XX detection and treatment of lung cancer. The present sequence is human

XX lung tumour-specific protein

XX Sequence 163 AA;

XX Alignment Scores:

XX Pred. No.: 3.75 Length: 163

XX Score: 86.00 Matches: 39

XX Percent Similarity: 43.22% Conservative: 12

XX Best Local Similarity: 33.05% Mismatches: 50

XX Query Match: 10.24% Indels: 17

XX DB: 4 Gaps: 5

XX US-10-057-510-1 (1-486) x AAE13757 (1-163)

XX QY 462 AGGCTTTCGNAACACACAGGGGNGTANGANCAGTTTCCAGACAGAGCCAGGACC 403

XX Db 11 ArgLeuSerLeuArgPro-----SerileProGluLeuSerGluArgThr 25

XX QY 402 ACAACAGGAGGGGGCTCTTATGCAAGTTCCTCATGCTATGATGAGCCAGGAGGACC 343

XX Db 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerProAla 45

QY 342 GTGGCTCGAACCTTGACTGGCAATGTCC-----TGAGAGTCTGGGGACNATCCA 292
 Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaGlnGluThrPro 65
 QY 291 GGGGCTGCATCTTCATCCCTGAGCCCTGGTGTAGCCTTCTGCGCTCTCCCTCCAGGGGCTGCTCC 232
 Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85
 QY 231 ATTCTTTCAGCTTCAGCGCCCTCCACACAGCTCCATCCCTCCACCTTGCCCATGGA-- 174
 Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyValGlyArgSerGlyAl 101
 QY 173 -NACAGCGGTGCCCC-----ACAAGCCAGAGGGAACAACAGAGCTGG 132
 Db 101 aAlaAlaIleProLeuGluLeuThrGlnSerArgValGlnIleTrp 118

RESULT 14
 ADD66349
 ID ADD66349 standard; protein; 163 AA.
 AC ADD66349;
 XX 15-JAN-2004 (first entry)
 XX Human lung tumour-specific protein sequence, SEQ ID NO 41.
 DE expression control; cancer; T cell; tumour; immune; cytostatic; vaccine;
 KW human; lung tumour-specific.
 XX Homo sapiens.
 XX WO200292001-A2.
 XX 21-NOV-2002.
 XX 10-MAY-2001; 2002WO-US014975.
 XX 11-MAY-2001; 2001US-00854133.
 XX (CORI-) CORIXA CORP.
 XX Lodes MJ, Wang T, Fan L, Algate PA, McNeill PD;
 WPI; 2003-120592/11.
 DR N-PSDB; ADD67083.
 XX
 PT New polynucleotide and polypeptide, useful for preparing a composition
 for diagnosing, treating or preventing cancer.
 XX
 PS Example 2; SEQ ID NO 41; 494pp; English.
 XX
 CC The invention relates to a novel isolated polynucleotide comprising one
 of 32 47-6080 base pair sequences, given in the specification, or their
 complements or degenerate variants, at least 20 contiguous residues of a
 sequence in, or having at least 75 or 90 % identity with the isolated
 polynucleotide, or that hybridise with the polynucleotide. The invention
 further comprises: an isolated polypeptide; an expression vector
 comprising the polynucleotide operably linked to an expression control
 sequence; a host cell transformed or transfected with the expression
 vector; an isolated antibody or its antigen-binding fragment that
 specifically binds to the polypeptide; a method for detecting the
 presence of a cancer in a patient; a fusion protein comprising the
 polypeptide; an oligonucleotide that hybridises to the isolated
 polynucleotide under moderately stringent conditions; a method for
 stimulating and/or expanding T cells specific for a tumour protein; an
 isolated T cell population; a composition comprising a first component
 consisting of carriers and immunostimulants and a second component; a
 method for stimulating an immune response in a patient; a method for
 treating cancer in a patient; a method for determining cancer in a
 patient; a diagnostic kit comprising at least one oligonucleotide or
 antibody and a detection reagent comprising a reporter group; and a
 method for inhibiting the development of cancer in a patient. The
 compositions of the invention have cytostatic activity and can be used to

CC create a vaccine. The isolated polynucleotide is useful for preparing a
 CC composition for diagnosing, treating or preventing cancer. This sequence
 CC represents a human lung tumour-specific protein relating to the
 CC invention.
 XX
 SQ Sequence 163 AA;
 Alignment Scores:
 Pred. No.: 3.75 Length: 163
 Score: 86.00 Matches: 39
 Percent Similarity: 43.22% Conservative: 12
 Best Local Similarity: 33.05% Mismatches: 50
 Query Match: 10.24% Indels: 17
 DB: Gaps: 5

US-10-057-510-1 (1-486) x ADD66349 (1-163)

QY 462 AGGCTTTGCNAAACACAGGGGNGCTANGANCAGATTTCCAGAACAGAGGGCCAGGACC 403
 Db 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluArgThr 25
 QY 402 ACAACAGGAGGGGGCTTCTTATGCAAGTTCCCTTCATGCCATATGAGCCCAAGNCCTN 343
 Db 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
 QY 342 GTGGCTCGAACCTTGACTGGCAATGTCC-----TGAGAGTCTGGGGACNATCCA 292
 Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaGlnGluThrPro 65
 QY 291 GGGGCTGCATCTTCATCCCTGAGCCCTGGTGTAGCCTTCTGCGCTCTCCCTCCAGGGGCTGCTCC 232
 Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85
 QY 231 ATTCTTTCAGCTTCAGCGCCCTCCACACAGCTCCATCCCTCCACCTTGCCCATGGA-- 174
 Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyValGlyArgSerGlyAl 101
 QY 173 -NACAGCGGTGCCCC-----ACAAGCCAGAGGGAACAACAGAGCTGG 132
 Db 101 aAlaAlaIleProLeuGluLeuThrGlnSerArgValGlnIleTrp 118

RESULT 15
 ADE87603
 ID ADE87603 standard; protein; 163 AA.
 XX
 AC ADE87603;
 XX 29-JAN-2004 (first entry)
 XX
 DE Human lung tumour antigen polypeptide #10.
 KW Human; lung tumour antigen; cancer; lung cancer; CD8+; T cell;
 KW immune response; immunostimulant; cytostatic.
 XX
 OS Homo sapiens.
 XX US2003118599-A1.
 XX 26-JUN-2003.
 XX 10-MAY-2002; 2002US-00144649.
 XX 02-APR-1999; 99US-00285323.
 XX 09-AUG-1999; 99US-00370838.
 XX 30-DEC-1999; 99US-00476235.
 XX 03-MAR-2000; 2000US-00518809.
 XX 29-MAR-2000; 2000US-00538037.
 XX 05-JUN-2000; 2000US-00588937.
 XX 18-AUG-2000; 2000US-00640878.
 XX 20-SEP-2000; 2000US-00667170.
 XX 01-NOV-2000; 2000US-00704512.
 XX 14-DEC-2000; 2000US-00738973.
 XX 11-MAY-2001; 2001US-00854133.

Job time : 59.5 secs

(CORI-) CORIXA CORP.

Algate PA, Lodes MJ, Wang T, Fan L, Mcneill PD;
WPI; 2003-897103/82.
N-PSDB; ADE87588.

New polynucleotides encode lung tumor antigens and are useful to stimulate an immune response or detect or treat a cancer in a patient, particularly lung cancer.

Example 2; SEQ ID NO 41; 63pp; English.

The invention relates to polynucleotides encoding lung tumour antigens. The invention also relates to the polypeptides encoded by the polynucleotides, isolated antibodies or antigen-binding fragments that specifically bind the polypeptides and a method for detecting cancer in a patient, comprising obtaining a biological sample from the patient, contacting the sample with a binding agent that binds a polypeptide of the invention, detecting in the sample an amount of polypeptide that binds to the binding agent, and comparing the amount of polypeptide to a predetermined cut-off value. T cells specific for a tumour protein can be stimulated and/or expanded by contacting the T cells with a polypeptide, polynucleotide or an antigen-presenting cell that expresses a polypeptide. Cancer development can be inhibited by incubating CD4+ and/or CD8+ T cells isolated from a patient with a polypeptide, polynucleotide or an antigen-presenting cell that expresses a polypeptide, so that the T cells proliferate. The invention is used to stimulate an immune response or to detect or treat a cancer in a patient, particularly lung cancer. This sequence represents a human lung tumour antigen polypeptide of the invention. Note: the sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 163 AA;

Alignment Scores:

Pred. No.:	3.75	Length:	163
Score:	86.00	Matches:	39
Percent Similarity:	43.22%	Conservative:	12
Best Local Similarity:	33.05%	Mismatches:	50
Query Match:	10.24%	Indels:	17
DB:	7	Gaps:	5

US-10-057-510-1 (1-486) x ADE87503 (1-163)

QY 462 AGGCTTTGCNAAANAACGAGGGGNGCTANGANCAGTTTTCAGAACAGAGGGCCAGACC 403

Db 11 ArgLeuSerLeuArgPro-----SerileProGluLeuSerGluArgThr 25

QY 402 ACACACGAGGGGGGCTCTTATGCAAGTTCCCTTCATGCCCATATGACGACAGNCCCTN 343

Db 26 SerArgProCyArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45

QY 342 GTGGCTCGAACCTTGACTGGGCAATGCC-----TGAGAGTCTGGGGACNATCCA 292

Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65

QY 291 GGGGCTGCATCTTCATCTCACCCTCGGTAGCCTCTCGCCCTCTCCAGGGCTCGTCC 232

Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyGlyArgSer 85

QY 231 ATTCCTTCAGCTCCAGCGCTCTCCACACGCTCATGCTTCCACCTTGCCACATGGA-- 174

Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101

QY 173 -NACAGCCGTGCCCC-----ACAGCCAGAGGGGAACAACAGACTGG 132

Db 101 aaAlaAlaValIleProLeuGluLeuThrGlnSerArgValGlnIleTrp 118